

HUMBOLDT-UNIVERSITÄT ZU BERLIN
INSTITUT FÜR BIBLIOTHEKS- UND INFORMATIONSWISSENSCHAFT



BERLINER HANDREICHUNGEN
ZUR BIBLIOTHEKS- UND
INFORMATIONSWISSENSCHAFT

HEFT 353

LANGZEITARCHIVIERUNG INNERHALB
VIRTUELLER FORSCHUNGSUMGEBUNGEN
IM BEREICH DIGITAL HUMANITIES

VON
SILKE BAUMANN

LANGZEITARCHIVIERUNG INNERHALB
VIRTUELLER FORSCHUNGSUMGEBUNGEN
IM BEREICH DIGITAL HUMANITIES

VON
SILKE BAUMANN

Berliner Handreichungen zur
Bibliotheks- und Informationswissenschaft

Begründet von Peter Zahn
Herausgegeben von
Konrad Umlauf
Humboldt-Universität zu Berlin

Heft 353

Baumann, Silke

Langzeitarchivierung innerhalb Virtueller Forschungsumgebungen im Bereich Digital Humanities / von Silke Baumann. - Berlin : Institut für Bibliotheks- und Informationswissenschaft der Humboldt-Universität zu Berlin, 2014. - 84 S. : graph. Darst. - (Berliner Handreichungen zur Bibliotheks- und Informationswissenschaft ; 353)

ISSN 14 38-76 62

Abstract:

In der vorliegenden Arbeit wird an einem konkreten Beispiel betrachtet werden, wie wissenschaftliche digitale Kooperation und die langfristige Speicherung der digitalen Daten im Bereich Digital Humanities aktuell umgesetzt werden bzw. welche Wünsche und Ansichten die ForscherInnen diesbezüglich haben. Zu diesem Zweck wurden WissenschaftlerInnen der Johannes Gutenberg-Universität Mainz zur Beantwortung eines Online-Fragebogens eingeladen. Einführend werden die schwer zu greifenden Begriffe Digital Humanities, Virtuelle Forschungsumgebung, Forschungsdaten und Langzeitarchivierung für den Kontext dieser Arbeit definiert. Es wird betrachtet, was prinzipiell die Anforderungen an Virtuelle Forschungsumgebungen, Forschungsdaten und Langzeitarchivierung im Bereich Geisteswissenschaften sind, wie der derzeitige Entwicklungsstand hier ist und welches Angebot eventuell schon existiert. Im Anschluss folgt ein Blick auf einige Vergleichsstudien, der den bisherigen Kenntnisstand zu Data Sharing und Langzeitarchivierung wiedergibt. Zusammen mit den wissenschaftlichen Hintergrundinformationen ist dieser auch Basis für die Hypothesen, die an den Ergebnissen der Befragung verifiziert werden. Nach Erläuterung der Methode und Methodologie der Datenerhebung folgen eine Beschreibung der erhobenen Daten und daraus resultierende Schlussfolgerungen.

Diese Veröffentlichung geht zurück auf eine Masterarbeit im postgradualen Fernstudiengang M. A. Bibliotheks- und Informationswissenschaft (Library and Information Science) an der Humboldt-Universität zu Berlin.

Online-Version: <http://edoc.hu-berlin.de/series/berliner-handreichungen/2014-353>



Dieses Werk steht unter einer Creative Commons [Namensnennung-NichtKommerziell-KeineBearbeitung 3.0](https://creativecommons.org/licenses/by-nc-nd/3.0/de/) Deutschland-Lizenz.

Inhalt

1. Einleitung.....	6
2. Wissenschaftlicher Kontext.....	10
2.1 Digital Humanities.....	10
2.2 Virtuelle Forschungsumgebungen.....	12
2.3 Forschungsdaten.....	18
2.4 Langzeitarchivierung.....	22
2.5 Vergleichsstudien.....	33
2.6 Hypothesen.....	40
3. Methodologie der Arbeit.....	41
4. Methode der Befragung.....	45
5. Ergebnisse der Befragung	48
5.1 Demographische Informationen.....	48
5.2 Art der digitalen Objekte.....	51
5.3 Metadaten.....	52
5.4 Langzeitarchivierung bzw.-verfügbarkeit der digitalen Objekte	53
5.5 Data Sharing.....	55
6. Fazit und Ausblick	58
7. Abbildungsverzeichnis.....	63
8. Tabellenverzeichnis.....	64
9. Literaturverzeichnis	65
10. Anhang.....	73
10.1 Fragebogen Online-Ansicht.....	73
10.2 Umfrage Logikdatei.....	77

1. Einleitung

In etwa den letzten fünfzehn bis zwanzig Jahren haben sich international die Informationsinfrastrukturen und der Forschungsalltag an Hochschulen stark verändert: „Der epochale Wandel in der Informations- und Kommunikationstechnologie (z.B. Digitalisierung, Webtechnologie) führt zu grundlegenden Veränderungen des wissenschaftlichen Arbeitens“ (Kommission Zukunft der Informationsinfrastruktur 2011, 13). Digitale Medien und damit auch digitale Forschungsdaten bzw. -veröffentlichungen sind auf dem Vormarsch. WissenschaftlerInnen erheben, durch technische Neuerungen angespornt, eine große Zahl von Forschungsdaten in digitaler Form, die sie dann mit (internationalen) Fachkollegen gemeinsam beforschen. Diese kollaborative und oft auch interdisziplinäre Arbeit findet idealerweise in sogenannten *Virtuellen Forschungsumgebungen* statt, die es den WissenschaftlerInnen erlauben, ort- und zeitungebunden gemeinsam an einer Forschungsfrage zu arbeiten.

Virtual Research Environments (VREs) sind ein Beispiel für integrierte Forschungsumgebungen, die über verteilte Ressourcen wie Speicherplatz und Rechenleistung hinausgehen und Menschen und Werkzeuge untereinander verknüpfen. VREs sind formelle Infrastrukturen, in der verteilte Ressourcen und Werkzeuge vereinigt werden, um einem speziellen Zweck oder einer speziellen Benutzergruppe zu dienen. Sie antworten auf den exponentiellen Wachstum von online verfügbaren digitalen Informationen, auf die permanente Zunahme von formell definierten Kollaborationen in teilweise sehr großen Benutzergruppen und auf die zunehmend einfachere Verfügbarkeit von Tools, mit denen Datensätze manipuliert und erkundet werden können (Aschenbrenner et al. 2007, 13).

Dies geschieht heute schon sehr effizient und organisiert in den Naturwissenschaften. Bekannte Beispiele hierfür sind z.B. der am *CERN (Conseil Européen pour la Recherche Nucléaire)* entwickelte *LHC (Large Hadron Collider)* oder das Teleskop *ALMA (Atacama Large Millimeter / submillimeter Array)* - das größte derzeit existierende astronomische Projekt.

Allerdings ist diese Entwicklung in den *Digital Humanities* noch nicht so verbreitet wie in den Naturwissenschaften (Borgman 2009, 5), obwohl sich auch hier, durch die Entwicklung zu einer digitalen Gesellschaft hin, neue Forschungsfragen und eine Fülle an multimedialen,

digitalen Forschungsdaten ergeben. Die für die GeisteswissenschaftlerInnen bisher eher ungewöhnliche gemeinschaftliche Forschungstätigkeit (ibid., 47) mit modernen Analysetools kann auch hier zu Synergieeffekten führen. Vorrangig arbeiten die GeisteswissenschaftlerInnen in diesem Kontext „mit einer unterschiedlichen Anzahl an digitalen Werkzeugen, beispielsweise bei der Zerlegung von größeren Textmengen in ihre Bestandteile, bei der Analyse von Verlinkung einzelner Begriffe mit digitalen Wörterbüchern, bei Annotationen und so weiter“ (Enquete-Kommission 2013, 83).

Durch die Bearbeitung von Forschungsfragen in einer Virtuellen Forschungsumgebung ergibt sich zukünftig auch die Möglichkeit einer Nachnutzbarkeit: zum einen können die entwickelten Module und Werkzeuge für weitere Forschungsprojekte weiterentwickelt und -genutzt werden, zum anderen ergibt sich durch integrierte Langzeitarchivierungslösungen die Möglichkeit, Forschungsdaten professionell zu sichern und für die Wissenschaftscommunity und für neue Forschungsfragen nachnutzbar zu halten. Derzeit steht man aber noch am Anfang dieser Entwicklungen, da noch viele technische, finanzielle und rechtliche Fragen in diesem Zusammenhang zu klären sind.

In der vorliegenden Arbeit soll an einem konkreten Beispiel betrachtet werden, wie wissenschaftliche digitale Kooperation und die langfristige Speicherung der digitalen Daten im Bereich *Digital Humanities* aktuell umgesetzt werden bzw. welche Wünsche und Ansichten die ForscherInnen diesbezüglich haben. Zu diesem Zweck wurden WissenschaftlerInnen der Fachbereiche Philosophie und Philologie sowie Geschichts- und Kulturwissenschaften der Johannes Gutenberg-Universität Mainz zur Beantwortung eines Online-Fragebogens eingeladen. Forschungsfragen waren: Welche Digitalen Daten bearbeiten die WissenschaftlerInnen? Ob und wie werden diese Daten mit Metadaten beschrieben? Ob und wie bearbeiten die Mainzer WissenschaftlerInnen ihre Forschungsdaten gemeinschaftlich mit Fachkollegen? Wie denken die WissenschaftlerInnen über die (längerfristige) Relevanz ihrer Forschungsdaten? Erfolgt eine Sicherung der erhobenen Forschungsdaten? Wie können sich die Mainzer WissenschaftlerInnen eine Langzeitarchivierung ihrer Daten vorstellen?

Im Folgenden werden zuerst die schwer zu greifenden Begriffe *Digital Humanities*, *Virtuelle Forschungsumgebung*, *Forschungsdaten* und *Langzeitarchivierung* für den Kontext dieser

Arbeit definiert. Es wird zu betrachten sein, was prinzipiell die Anforderungen an Virtuelle Forschungsumgebungen, Forschungsdaten und Langzeitarchivierung im Bereich Geisteswissenschaften sind, wie der derzeitige Entwicklungsstand hier ist und welches Angebot eventuell schon existiert. Im Anschluss folgt ein Blick auf einige Vergleichsstudien, der den bisherigen Kenntnisstand zu Data Sharing und Langzeitarchivierung wiedergibt. Zusammen mit den wissenschaftlichen Hintergrundinformationen ist dieser auch Basis für die Hypothesen, die an den Ergebnissen der Befragung verifiziert werden sollen. Nach Erläuterung der Methode und Methodologie der Datenerhebung folgen eine Beschreibung der erhobenen Daten und daraus resultierende Schlussfolgerungen.

2. Wissenschaftlicher Kontext

2.1 Digital Humanities

Untersucht man das Thema „Langzeitarchivierung in Virtuellen Forschungsumgebungen im Bereich Digital Humanities“ (oder auch e-Humanities) ist es unumgänglich, den Begriff *Digital Humanities* zu definieren. Dies ist ein extrem schwieriges, aber oft unternommenes Unterfangen und hat, wie Matthew G. Kirschenbaum in 2010 feststellte, darin resultiert, dass diese Aufsätze bereits als „genre pieces“ (Kirschenbaum 2010, 1) bezeichnet werden können. Allerdings gehen die Meinungen darüber, was *Digital Humanities* überhaupt heißt, sehr weit auseinander.

Seit einigen Jahren wird der *Day of Digital Humanities* veranstaltet, ein gemeinschaftliches Projekt von WissenschaftlerInnen auf der ganzen Welt, die einen Tag ihrer Forschung dokumentieren. Auf dem Anmeldungsformular zu diesem Ereignis werden die TeilnehmerInnen gebeten, ihre persönliche Definition der *Digital Humanities* zu formulieren. Die Begriffserklärungen der Personen, die ihr Einverständnis dazu erteilten, sind im *Taporwiki* der Universität von Alberta veröffentlicht worden und so verschieden wie zahlreich. Hier nur einige Beispiele:

- Developing tools and workflows to create comprehensive, interoperable, and innovative digital resources. – *Jennifer Stertz, Papers of George Washington, University of Virginia, USA*
- Humanities research based on Digital Data and the Internet. – *Kozaburo Hachimura, Ritsumeikan University, Japan*
- Humanities by digital means. *Franz Fischer, Royal Irish Academy, Ireland*
- A 'community of practice' (to borrow Etienne Wenger's phrase) whereby the learning, construction and sharing of humanities knowledge is undertaken with the application of digital technologies in a reflexive, theoretically-informed, and collaborative manner. – *Kathryn E. Piquette, Humboldt-Universität, Germany*
- Digital Humanities for me is the way that we can collaborate and share online, all over the world, bringing to life again the old stories while creating new stories of our own, in words in images. – *Laura Gibbs, University of Oklahoma, USA*

- The digital humanities study how digital media (formats and tools) and the cultural forms associated with them affect the work of the humanities – and vice versa. – *Lewis Ulman, The Ohio state University, USA*
- I don't define Digital Humanities any more. All humanities seem to be digital in one way or another. – *Martin Holmes, University of Victoria, Canada*

Der Begriff *Digital Humanities* hat sich aus den *Humanities Computing* entwickelt (Kirschenbaum 2010, 2). Roberto Busa, ein italienischer Ordensgeistlicher und Linguist, gilt als der Begründer der wissenschaftlichen Anwendung der EDV in den Geisteswissenschaften. Er führte, zusammen mit Thomas J. Watson, dem Gründer von IBM, die computergestützte Lemmatisierung der Werke von Thomas von Aquin, dem *Index Thomisticus*, durch (Hockey 2004). Er begann dieses monumentale Unterfangen bereits 1949 und war 1974 soweit den ersten gedruckten Band herauszubringen. Seither wurden die digitalen Geisteswissenschaften mit der Entwicklung und Anwendung computergestützter Verfahren für die Analyse von Texten (z.B. Text Encoding Initiative) oder auch mit den Aufbau der systematischen Verwendung von digitalen Ressourcen (z.B. William Blake Archive) in den Kultur-, Geistes-, und Sozialwissenschaften assoziiert (Fitzpatrick 2012). Bislang hat sich die digitale Geisteswissenschaft in eine Disziplin mit immer diffuserer Beschreibung entwickelt, wie auch der, Mitte 2012 gegründete, Verband *DHd - Digital Humanities im deutschsprachigen Raum*, auf seiner Homepage verdeutlicht:

Manche sehen in ihnen [Digital Humanities] eine Hilfswissenschaft zu den geisteswissenschaftlichen Fächern, deren Fragestellungen sie aufgreifen. Manche betrachten sie als eine angewandte Informatik, die für einen speziellen Gegenstandsbereich Informatik-Lösungen entwickeln. Wieder andere sehen in ihr eine allumfassende Geisteswissenschaft auf der Ebene der Methoden und digitalen Praktiken.

Es ist im Rahmen dieser kurzen Arbeit leider nicht möglich, in eine vertiefende Diskussion der verschiedenen Ausprägungen des Begriffes einzusteigen, daher muss eine sehr weitgefasste und vereinfachte Definition genügen.

Es ist festzustellen, dass die digitale Welt heutzutage fast alle Bereiche des menschlichen Lebens durchdringt, so auch die Welt der Forschung. WissenschaftlerInnen benutzen „word processors instead of pencils, computers instead of card catalogs, text encoding instead of

notes scribbled in margins and notecards" (Parry 2012). Insofern ist es möglich zu behaupten, dass „[t]he practice of the humanities, of the academy as a whole (certainly within the American and European contexts), is thoroughly integrated with the digital and is, at this point, impossible to separate from it" (ibid.). *DHers* mögen es verzeihen, wenn daher für den Kontext dieser Arbeit, *Digital Humanities* als Wissenschaft in den Geistes-, Kultur-, und Geschichtswissenschaften so wie sie 2013 in ihrer digitalen Vielfalt praktiziert wird, gemeint sein soll oder wie David Golumbia es in seinem Blog formuliert: „anything that combines digital work of any sort with humanities work of any sort“(Golumbia 2013).

2.2 Virtuelle Forschungsumgebungen

Was genau ist eine Virtuelle Forschungsumgebung (oder auch Virtual Research Environment bzw. Cyberinfrastruktur)? Liest man sich in die einschlägige Fachliteratur ein, wird schnell klar, dass man ein ähnlich weites Feld wie bei den Begriffsbestimmungsversuchen der *Digital Humanities* betritt (Dickmann und Sax 2010, 4). Man ist sich immerhin einig, dass sie einen Teil der modernen Forschungsinfrastruktur konstituieren. Die *Kommission Zukunft der Informationsinfrastruktur (KII)*, die im April 2011 ihr *Gesamtkonzept für die Informationsinfrastruktur in Deutschland* vorlegte, definiert die wissenschaftliche Informationsinfrastruktur als „Bestandteil der Forschungsinfrastruktur“ (Kommission Zukunft der Informationsinfrastruktur 2011, 6), und führt weiter aus, dass „**Kernaufgabe** der Informationsinfrastruktur ist – im weitesten Sinne – die Versorgung von Wissenschaft und Forschung mit Information und damit zusammenhängenden Dienstleistungen und Diensten“ (ibid., 14-15). Dies beinhaltet: Erwerb, Aufbereitung, Erschließung, Nachweis, Bereitstellung und Archivierung von Informationen; Sicherstellung der nachhaltigen Retrieval- und Analysefähigkeit relevanter Information; Management von Information aller Art inklusive der Bereitstellung von Tools zur Bearbeitung; Sicherstellung des dauerhaften Zugriffs; Gewährleistung von Sicherheit, Integrität und Authentizität; Bereitstellung von Möglichkeiten der gemeinschaftlichen Nutzung und der elektronischen Kommunikation sowie Support dieser neuen Prozesse und Arbeitsgebiete durch geeignete Methoden in der Lehre und Ausbildung (ibid., 14). Dieses Konglomerat an Dienstleistungen und Diensten bedingt

idealerweise, dass die moderne Informationsinfrastruktur die komplette Wertschöpfungskette des wissenschaftlichen Forschungsprozesses unterstützt, beginnend bei der Formulierung der Forschungsfrage und der Projektplanung über die Datenerhebung und Recherche, die zu dem Austausch bzw. der Diskussion von Daten, Gedanken und Ergebnissen führen bis hin zu einer fundierten Analyse in der Publikation und deren Langzeitverfügbarkeit. Die publizierten Ergebnisse wiederum bieten neuen Stoff für die Lehre und geben Anreiz zu innovativen Forschungsfragen.



Abbildung 1: Der wissenschaftliche Wertschöpfungszyklus

Dieser gesamte Forschungszyklus kann von einer Virtuellen Forschungsumgebung unterstützt werden, die projektspezifische Daten und (international) verteilte Fachkenntnisse bündelt bzw. managt, kollaborative Arbeit und damit Interdisziplinarität und Internationalität fördert sowie dem Forschungsvorhaben angepasste state-of-the-art Analysetools zur Verfügung stellt – alles innerhalb einer Plattform. „One characteristic of infrastructure is that it is deeply imbedded in the way we work. When it works efficiently, it is invisible: we use it without really thinking about it“ (American Council of Learned Societies 2006, 6).

Die Virtuelle Forschungsumgebung stellt eine flexible One-Stop-Lösung dar, die durch Synergieeffekte eine Beschleunigung und Qualitätsverbesserung der Forschung bewirken kann. Allerdings ist es zwingend notwendig, dass sie zusammen mit den WissenschaftlerInnen entwickelt wird, da selbst die beste Plattform der Wissenschaft nur hilft, sofern sie von den ForscherInnen auch angenommen wird und einfach zu nutzen ist (Kommission Zukunft der Informationsstruktur 2011, B82). Gerade die Frage der einfachen *Usability* dürfte besonders für die Geisteswissenschaften ein wichtiger Punkt sein, da sie sich, im Kontrast zu den Naturwissenschaften, und aufgrund der noch recht weit verbreiteten Forschungskultur des Monographien veröffentlichenden Einzelkämpfers (American Council of Learned Societies 2006, 21), oft noch nicht mit den Cyberinfrastrukturen arrangiert haben.

Die Schwerpunktinitiative „Digitale Information“ der Allianz der deutschen Wissenschaftsorganisationen hat Virtuelle Forschungsumgebungen zu einem seiner sechs Handlungsfelder erklärt und definiert diese im Februar 2011 auf ihrer Homepage wie folgt:

Eine virtuelle Forschungsumgebung ist eine Arbeitsplattform, die eine kooperative Forschungstätigkeit durch mehrere Wissenschaftler an unterschiedlichen Orten zu gleicher Zeit ohne Einschränkungen ermöglicht. Inhaltlich unterstützt sie potentiell den gesamten Forschungsprozess – von der Erhebung, der Diskussion und weiteren Bearbeitung der Daten bis zur Publikation der Ergebnisse – während sie technologisch vor allem auf Softwarediensten und Kommunikationsnetzwerken basiert. Virtuelle Forschungsumgebungen sind wesentliche Komponenten moderner Forschungsinfrastrukturen.

Der größte Vorteil ist, dass WissenschaftlerInnen ganz unterschiedlicher Disziplinen zeit- und ortsunabhängig zusammenarbeiten können, wobei sie organisatorische und technische Unterstützung von Bibliothekaren und Informatikern erhalten. Die ForscherInnen haben sowohl Zugriff auf fachspezifische als auch auf interdisziplinäre Services, auf die sie ohne Medienbruch zugreifen können. Die Tools zur Bearbeitung und Beforschung der Materialien werden idealerweise von den ForscherInnen selbst mitentwickelt und können so perfekt den Forschungsbedürfnissen angepasst werden. Das gemeinsame Bearbeiten von Forschungsdaten sowie der generelle Austausch von Daten und Ergebnissen rücken immer mehr in den Vordergrund des wissenschaftlichen Arbeitens, was dazu beitragen kann, Wissenschaft transparenter, innovativer und kommunikativer zu gestalten.

Die *DINI Arbeitsgruppe Virtuelle Forschungsumgebungen* setzt so auch auf ihrer Homepage klar den Akzent bei den aktiv Forschenden und ihren Anforderungen an eine Forschungsinfrastruktur:

Der Begriff "Virtuelle Forschungsumgebung" beschreibt dabei ein breites Spektrum an IT-gestützten Forschungsszenarien, das von der Kommunikation in der Gruppe über das verteilte Schreiben bis zur gemeinsamen Durchführung von Experimenten und den Zugriff auf tausende Kilometer entfernte Geräte und Sensoren reicht. Über die wissenschaftlichen Inhalte hinaus spielen vor allem die Aspekte „wie“ (arbeiten Wissenschaftlerinnen und Wissenschaftler über das Internet zusammen) und „was“ (brauchen sie dazu) eine zentrale Rolle bei der Entwicklung von IuK-Dienstleistungen für Virtuelle Forschungsumgebungen.

Schnell wird deutlich, dass es aber auch noch viel zu tun gibt, um diese innovativen Forschungsräume tatsächlich international optimal verwirklichen zu können. Es muss also umfassend geklärt werden, wie die GeisteswissenschaftlerInnen bisher interagieren und wie sie sich eine Optimierung ihres Forschungsalltags vorstellen. Weiterhin muss klar sein, was die Geisteswissenschaften eigentlich an Hilfestellungen und Werkzeugen benötigen, um ihre Forschungsfragen zielbringend verfolgen zu können. Hierzu sind umfassende Befragungen notwendig, wozu der im Rahmen dieser Masterarbeit erfolgte Internet-Survey einen kleinen Beitrag leisten kann.

Des Weiteren bedarf es der Entwicklung international einheitlicher Standards für die Datenbeschreibung, d.h. einheitliche Metadatenstandards zwecks Interoperabilität der Systeme. Dafür müssen Format- und Softwarestandards entwickelt werden. Im Prinzip muss die gesamte Datenverwaltung (*Data Curation*) noch auf eine gesunde Basis gestellt werden.

Derzeit ist die Situation noch geprägt durch mangelnde Transparenz über Verfügbarkeit und Qualität, fehlende Metadaten und Standardisierung, Das sogenannte Lebenszyklusmanagement ebenso wie der rechtliche sowie der ethische Umgang mit Forschungsdaten stellen weitere Herausforderungen dar (Kommission Zukunft der Informationsstruktur 2011, 17).

Es müssen offene Schnittstellen bereitgestellt werden, die den offenen Datenaustausch ermöglichen. Die Infrastruktursysteme sollten modular gestaltet werden, damit sie an die

sich ändernden Forschungsanliegen angepasst werden können und bestehende Ressourcen auch wiederverwenden können. Die Frage der Langzeitarchivierung muss adressiert werden, damit die gewonnenen Forschungsdaten und -ergebnisse nachnutzbar bleiben. Auch im Kontext der Langzeitarchivierung sind Fragen technischer sowie auch rechtlicher Natur, wie z.B. das Urheberrecht bezüglich Forschungsdaten und gemeinschaftlich erarbeiteter Inhalte, sowohl im nationalen als auch im internationalen Kontext noch zu klären. Es gilt, Virtuelle Forschungsumgebungen aus dem Projektbetrieb in einen langfristigen, finanziell abgesicherten Dauerbetrieb zu überführen. Schließlich müssen die WissenschaftlerInnen noch im Umgang mit den neuen Daten, Standards und Werkzeugen ausgebildet werden (ibid., B73 ff.).

Es muss also noch viel passieren ... es ist allerdings auch schon sehr viel geschehen. An dieser Stelle sollen nur zwei bereits bestehende wichtige Forschungsinfrastrukturen im Bereich *Digital Humanities* Erwähnung finden. Zum einen ist da die im Rahmen des *European Strategic Forum on Research Infrastructures (ESFRI)* geförderte *DARIAH-EU*. „The mission of DARIAH is to enhance and support digitally-enabled research across the humanities and arts. DARIAH aims to develop and maintain an infrastructure in support of ICT-based research practices“ (DARIAH-EU, Homepage). *DARIAH-EU* soll im Bereich der Geisteswissenschaften dazu beitragen, die Entwicklung des Forschungsstandortes Europa bis zum Jahre 2020 auf ein Spitzenniveau zu bringen, wie in der *Roadmap* des ESFRI ausführlich beschrieben wird (2010, 24). Daneben gibt es auch noch eine nationale, vom Bundesministerium für Bildung und Forschung (BMBF) geförderte Projektgruppe, *DARIAH-DE*, die ihren Auftrag darin sieht:

[...] die Interoperabilität von Werkzeugen und Forschungsdaten sowie deren Langzeitarchivierung und Nachnutzung zu ermöglichen. Hierbei orientiert sich DARIAH-DE an international gültigen und akzeptierten Standards und Vorgehensweisen und engagiert sich bei deren Weiterentwicklung. Um den Mehrwert von Methoden, Diensten und Werkzeugen der Digital Humanities anschaulich zu machen, werden beispielhafte Anwendungsprojekte identifiziert. Basierend darauf werden einzelne Lösungen als sogenannte Demonstratoren ausgewählt und gegebenenfalls weiter entwickelt, um Möglichkeiten und Arbeitsweisen einer digitalen Forschungsinfrastruktur im prototypischen Betrieb zu zeigen (DARIAH-DE, Projektziele).

Ebenfalls vom BMBF gefördert wird das seit 2006 bestehende Verbundprojekt *TextGrid*. Hierbei handelt es sich um eine modulare Plattform für die textbasierte, gemeinschaftliche Forschung zur Erarbeitung digitaler Editionsprojekte.

Im Sommer 2011 stand die erste stabile Version 1.0 zum kostenfreien Download bereit, im Mai 2012 die Version 2.0. Schwerpunkt der gegenwärtigen Förderphase (2012-2015) ist schließlich die Etablierung eines nachhaltigen Dauerbetriebs: Software, Archiv und damit das gesamte Angebot sollen in Zusammenarbeit mit AnwenderInnen, FachwissenschaftlerInnen und Fachgesellschaften auf eine langfristige Nutzerbasis gestellt werden. (TextGrid, Das Projekt).

TextGrid ist das erste geisteswissenschaftliche Grid-Projekt in Deutschland und Teil der ebenfalls vom BMBF geförderten *D-Grid-Initiative* zum Aufbau einer Grid-Computing-Infrastruktur.

Am einfachsten versteht man Grids, vergleicht man es mit bestehenden Netzwerktechnologien wie dem Internet. Dieses kann als globales virtuelles Multimedia-Dokument betrachtet werden, von dem man einzelne Seiten mit seinem lokalen Browser herunterladen kann. All die anderen Ressourcen der Computer im Internet bleiben dabei ungenutzt (z.B. Prozessorleistung). Grids wollen diese ungenutzten Ressourcen in so genannten ‚Virtual Organisations‘ verfügbar machen, die Individuen und Institutionen jenseits von administrativen Grenzen in gemeinsamen Ressourcennutzungsrichtlinien zusammenfassen. [...] ‚On demand‘ stehen damit für den Wissenschaftler Kapazitäten bereit, die er auf seinem Rechner oder seinem Institutsnetz nicht zur Verfügung hätte (Aschenbrenner et al. 2007, 12).

Die *TextGrid* Infrastruktur besteht im Wesentlichen aus dem *TextGrid Respository*, einem fachwissenschaftlichen Archiv, in dem Forschungsdaten langfristig abgelegt werden können, und dem *TextGridLab*, welches unterschiedliche Open-Source Dienste und Tools in einer Oberfläche kombiniert. Das Lab ist modular gestaltet, sodass es beliebig ausgebaut und an das Forschungsinteresse angepasst werden kann (TextGrid, Tools Services Ressourcen).

2.3 Forschungsdaten

Wie bereits im vorherigen Kapitel erwähnt wurde, kommt den Forschungsdaten als Basis wissenschaftlicher Forschung eine immer größere Bedeutung zu. Schon die 2003 von den führenden deutschen Wissenschaftsorganisationen unterzeichnete *Berliner Erklärung über offenen Zugang zu wissenschaftlichem Wissen*, einer der wichtigsten Meilensteine der Open-Access Bewegung, stellt fest, dass auch Forschungsdaten frei zur Verfügung gestellt werden sollten:

Beiträge nach dem „Prinzip des offenen Zugangs“ können sowohl wissenschaftliche Forschungsergebnisse als auch Rohmaterialien und Metadaten, Quellenmaterialien, digitale Repräsentationen von bildlichem und grafischem Material sowie wissenschaftliche Materialien in multimedialer Form sein (Berliner Erklärung über offenen Zugang zu wissenschaftlichem Wissen 2003).

Für großes Aufsehen sorgten auch die 2007 veröffentlichten *OECD Principles and Guidelines for Access to Research Data from Public Funding*, welche fordern, dass Forschungsdaten, die mit öffentlichen Mitteln gefördert wurden offen zugänglich gemacht werden sollten: „Sharing and open access to publicly funded research data not only helps to maximize the research potential of new digital technologies and networks, but provides greater returns from the public investment in research“(OECD Principles and Guidelines for Access to Research Data from Public Funding 2007, 10).

Die WissenschaftlerInnen nehmen in diesem Kontext eine duale Rolle ein, da sie gleichzeitig als Produzenten sowie auch Konsumenten des intellektuellen „Rohstoffs“ agieren. Daher sind auch die Forschungsdaten eines der sechs Handlungsfelder der *Allianzinitiative „Digitale Information“*. Sie definiert Forschungsdaten als „Daten, die im Zuge wissenschaftlicher Vorhaben z.B. durch Digitalisierung, Quellenforschung, Experimente, Messungen, Erhebungen oder Befragungen entstehen“ (Allianz der deutschen Wissenschaftsorganisationen, Forschungsdaten).

Sowohl die *Allianzinitiative* als auch die *Kommission Zukunft der Informationsinfrastruktur*, die *DFG* sowie *DINI (AG Kompetenzzentrum Interoperable Metadaten / KIM)* haben im Zusammenhang mit Forschungsdaten allerdings noch ausdrücklich einigen Handlungsbedarf

bezüglich der „Sicherung, Erschließung, Bereitstellung und langfristigen Bewahrung von Forschungsdaten“ (Kommission Zukunft der Informationsinfrastruktur 2011, 43) identifiziert. Die *Arbeitsgruppe Forschungsdaten* des KIM formuliert es in ihrem Wiki folgendermaßen:

Forschungsdaten sind kein Nebenprodukt der Forschung. Vielmehr bilden sie eine wertvolle Wissensressource, welche durch professionelles Datenmanagement dauerhaft gesichert und leicht zugänglich gemacht werden muss. Sichern, Archivieren, Teilen und Nachnutzen von Forschungsdaten bringen Kostenersparnis und Mehrwert für die Wissenschaft.

Das Datenmanagement, welches einen erheblichen Arbeitsaufwand erfordert, und hier insbesondere die Anforderungen an die Beschreibung der Daten durch standardisierte Metadaten, müssen noch genau erarbeitet werden (Higgins 2007, What are Metadata Standards).

Doch was genau sind eigentlich Metadaten? Metadaten sind geordnete Daten, die Informationen über Informationsobjekte oder -ressourcen beinhalten und diese dadurch beschreiben und dokumentieren. Diese Informationsobjekte können Texte, Bilder, Filme, Kunstwerke, historische Artefakte, Klänge, Webseiten – kurz, alle Arten von Objekten – sein. Diese Daten über Daten können, je nach Forschungsdisziplin, Informationen über administrative, technische oder strukturelle Charakteristika von Ressourcen enthalten.

Metadaten sind Daten oder Informationen, die in strukturierter Form analoge oder digitale Forschungsdaten (Objekte) dokumentieren. Sie beschreiben, erklären, verorten oder definieren Objekte, Ressourcen und Informationsquellen für die Wissenschaft. Hierdurch helfen sie, Forschungsdaten zu managen, zu erschließen, zu verstehen und zu benutzen (Jensen, Katsanidou und Zenk-Möltgen 2011, 83).

Es haben sich bislang einige Standards zur Indexierung, also dem Schreiben von Metadaten, etabliert, die allerdings noch lange nicht so verbreitet sind, wie das im Sinne einer wirklichen Interoperabilität von allen Systemen wünschenswert wäre. Interoperabilität meint hier, dass die Metadaten verschiedenen Ursprungs so „gebaut“ sind, dass man trotzdem mit ihnen gleichzeitig arbeiten kann. Zur Erreichung dieser Interoperabilität von Metadaten müssen Standards erstellt werden, die sich auf Semantik, Syntax, Datenmodell und Identifikation der Metadaten konzentrieren. Einige bereits etablierte Standards sind z.B. *Dublin Core Metadata*

Element Set (ISO Standard 15836), *PREMIS* (Data Dictionary for Metadata Preservation), *DDI* (Data Documentation Initiative) oder *RDF* (Resource Description Framework).

Auch gibt es verschiedenen Arten von Metadaten, die unterschiedliche Funktionen in Bezug auf eine Informationsressource haben wie bei Harvey genau erläutert wird:

- *Descriptive metadata*, which describes digital objects in ways that ensure that they can be identified and located;
- *Structural metadata*, which indicates the relationships between one digital object and others;
- *Technical metadata*, which records details of technical aspects of the creation, encoding and formatting of digital objects, which are required so that these objects can be used;
- *Administrative metadata*, which describes the process applied to digital objects over time and may include rights management metadata and other kinds of metadata;
- *Preservation metadata* (sometimes considered as a subset of administrative metadata), which records the provenance of digital objects and the actions applied to them as they are managed over time in a digital archive (Harvey 2012, 83-84).

Ganz besonders wichtig wird die genaue Beschreibung der Ressourcen mit Hinblick auf die Langzeitverfügbarkeit; wie im nächsten Kapitel zu sehen sein wird.

WissenschaftlerInnen sollten also im Umgang mit Metadaten geschult sein und ein vitales Interesse daran haben, dass ihre Forschungsdaten vollständig und interoperabel beschrieben sowie sicher und langfristig verwaltet und gespeichert werden. Die Realität sieht allerdings anders aus: „Jahrzehnte lang dominierte aber die Ablage in lokalen Systemen, die Datenspeicherung und -weitergabe wurde auf der Ebene von Arbeitsgruppen oder Instituten praktiziert“ (Büttner et al. 2011, 18). Die WissenschaftlerInnen haben die Daten oft ganz ohne Beschreibung oder zumindest ohne Beachtung international anerkannter Normen, ohne Unterstützung von Informationsprofis, verwaltet und gespeichert.

Es ist auch bisher eine allgemeine Abneigung der WissenschaftlerInnen bezüglich der Freigabe ihrer Daten festzustellen: „Scholars compete as well as collaborate, and thus have reasons *not* to share data sources“ (Borgman 2009, 54). Neben dem ausgeprägten

Konkurrenzdenken bei ForscherInnen identifiziert der *Wissenschaftsrat* auch noch technische, normative und finanzielle Hindernisse, Forschungsdaten zu teilen (Empfehlungen zu den Forschungsinfrastrukturen in den Geistes- und Sozialwissenschaften 2011, 55), obwohl mittlerweile als nachgewiesen gilt, „dass das Teilen von Daten mit höheren Zitationsraten verbunden ist“ (Huschka et al. 2011, 39).

Ein „verantwortungsvoll[er] und organisiert[er] Umgang mit Forschungsdaten“ (Rümpel 2011, 25) beinhaltet, dass ihr gesamter Lebenszyklus professionell verwaltet wird. Ein guter Anhaltspunkt zur Identifikation aller Stufen des Forschungsdatenmanagements ist das *Curation Lifecycle Model* des *Digital Curation Centre* (DCC, 2010).

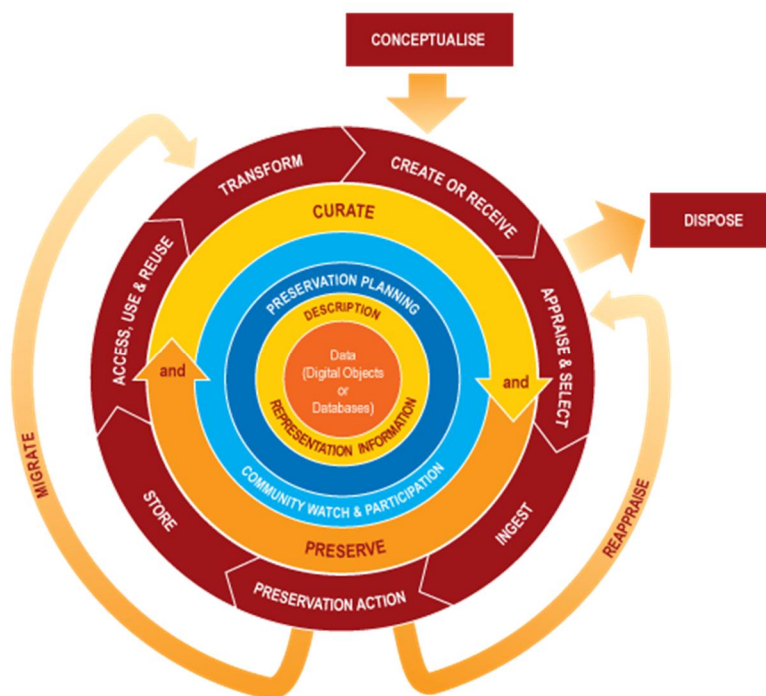


Abbildung 2: Curation Lifecycle Model (DCC 2010)

Im Zentrum des Datenverwaltungs-Modells befinden sich die Daten, also jegliche Art von Information in binärer, digitaler Form. Dies können digitale Objekte sein, wie z.B. Text-Dateien oder Bild-Dateien mitsamt ihren zugehörigen Metadaten und Identifikatoren. Auch Datenbanken, also strukturierte Sammlungen von digitalen Daten, sind hiermit gemeint. Um den Kern reihen sich in ringförmiger Anordnung die verschiedenen Aufgaben, die bei der Verwaltung von Daten anfallen, wobei sich zwei wesentliche Bereiche herausbilden:

„*Data Preservation*“ (Datenerhaltung) und „*Data Curation*“ (Datenpflege). Beide ergänzen sich und bilden die Kernprozesse der *Digital Curation*. Diese Arbeiten müssen im gesamten Lebenszyklus von Forschungsdaten erfolgen. *Preservation* bezieht sich auf die Bewahrung der Daten im Sinne der digitalen Langzeitarchivierung. Um Daten nutzbar zu gestalten und zu behalten, wird eine Pflege notwendig, subsumiert unter dem Begriff „*Data Curation*“ (Rümpel 2011, 28).

Das *DCC Curation Lifecycle Model* unterstützt die Benutzer, die Verwaltung der Daten zu planen und organisieren, sodass alle notwendigen Schritte in der richtigen Reihenfolge durchgeführt werden können. Gleichzeitig ist das Modell auf alle Fachdisziplinen anwendbar. Es hilft, die Verantwortlichkeiten zuzuordnen und bietet somit einen Rahmen an Standards und Technologien, die zur Verwaltung und Pflege von Forschungsdaten notwendig sind. Anhand des Modells ist es auch möglich, noch eventuell fehlende Teilschritte oder -aktionen zu identifizieren und einzubauen (DCC 2010).

Es ist sinnvoll, dieses Modell in die Struktur einer Virtuellen Forschungsumgebung einzubauen, um sicher zu gehen, dass der Forschungsprozess und die damit untrennbar verbundene Forschungsdatenverwaltung ideal organisiert werden. Die WissenschaftlerInnen könnten hierdurch zusätzlich unterstützt werden, da bei den meisten Forschungsanträgen derzeit ein detaillierter Forschungsdatenmanagementplan miteinzureichen ist.

2.4 Langzeitarchivierung

Leider gibt es bisher keine international anerkannte Definition von Langzeitarchivierung (Oßwald et al. 2012, 15). Es geht aber im Kontext der Forschung darum, Forschungsinhalte in Form von Forschungsdaten und Publikationen langfristig digital zur Verfügung stellen zu können, wobei die Inhalte verifizierbar, nachnutzbar und interpretierbar sein müssen, um für zukünftige Forschungsvorhaben von Nutzen zu sein (ibid., 15). Dabei können die Daten als *physisches Objekt*, als *logisches Objekt* und als *konzeptuelles Objekt* gespeichert werden (Thibodeau 2002).

„As a physical object, a digital object is simply an inscription of signs on a medium“ (ibid.). Es handelt sich also bei der Speicherung des *physischen Objekts* um einfache Speicherung der *Bits* und *Bytes* auf einem Datenträger ohne eine nähere Beschreibung des Objekts durch Morphologie, Syntax, oder Semantik (Metadaten). Bei dieser *Bitstream-Preservation* (oder Bitstrom-Speicherung) sollten auf Anraten des Kooperationsverbundes *nestor* – *Kompetenznetzwerk Langzeitarchivierung und Langzeitverfügbarkeit digitaler Ressourcen für Deutschland* folgende Punkte dringend beachtet werden:

1. *Redundante Datenhaltung*: Die Daten sollten in mehrfacher Kopie vorliegen. Zur Sicherung gegen äußere Einflüsse empfiehlt sich auch eine räumlich getrennte Aufbewahrung der unterschiedlichen Kopien.
2. *Diversität eingesetzter Speichertechnik*: Die Daten sollten auf mindestens zwei unterschiedlichen Datenträgertypen gesichert werden.
3. *Standards*: Die verwendeten Speichermedien sollten internationalen Standards entsprechen und auf dem Markt eine weite Verbreitung aufweisen.
4. *Regelmäßige Medienmigration*: Die verwendeten Speichertechniken bzw. Datenträger müssen regelmäßig durch neue ersetzt werden (Ulrich 2010, 8.2.).

Die *Bitstream-Preservation* ist Basis der Langzeitarchivierung und es ist wichtig, weitere Maßnahmen zur Langzeitspeicherung zu unternehmen, um die Daten interpretierbar und nachnutzbar zu machen.

„A digital information object is a logical object according to the logic of some application software“ (Thibodeau 2002). Die Folge von Bits und Bytes, die ein Dateiformat herausbilden, definieren das *logische Objekt*. Eine Software erkennt den Datenfluss und welches Format (z.B. PDF-Format) er repräsentiert. Es ist aber zwingend notwendig, dass das Format bekannt ist und die zugehörige Software zur Identifikation noch existiert, da sonst das Objekt nicht mehr rekonstruiert werden kann. Es gibt zwei wesentliche Strategien um diesem Problem zu begegnen: *Emulation* und *Migration*, auf die später noch kurz näher eingegangen werden soll.

„The conceptual object is the object we deal with in the real world: it is an entity we would recognize as a meaningful unit of information, such as a book, a contract, a map, or a photograph“ (Thibodeau 2002). Das *konzeptuelle Objekt* schließlich umfasst die gesamte Funktionalität, die der Benutzer durch Einsatz der richtigen Hard- und Software ausführen kann. Es ist der eigentliche Gegenstand, der durch Langzeitarchivierungsmaßnahmen erhalten werden soll:

Das Ziel eines Langzeitarchivs ist es also, das konzeptuelle Objekt zu archivieren und dem Nutzer auch in ferner Zukunft Zugriff auf dessen Inhalte zu gewähren. Die Darstellung bzw. Nutzung des digitalen Objekts soll so nahe wie möglich den Originalzustand des Objekts zur Zeit der Archivierung widerspiegeln. Dies ist nicht möglich, wenn sich bereits Probleme bei der Archivierung auf den unteren Ebenen, der logischen und der physischen Ebene, ergeben (Funk 2010, 7.6).

Es ist bisher also klar geworden, dass ohne einen unbeschädigten *Bitstream* und eine entsprechende Hard- und Software-Umgebung die Daten nicht wieder für den Nutzer lesbar gemacht werden können. Die *Bitstream-Preservation* muss durch die bereits erwähnten Methoden, also der redundanten Datenverwaltung, der Diversität von eingesetzter Speichertechnik, der Einhaltung von Standards und durch regelmäßige Medienmigration sichergestellt werden. Hierauf bauen nun *Migration* und *Emulation* auf, damit Dateninhalte langfristig nachnutzbar und interpretierbar bleiben. Die *Migration* assimiliert die Daten in eine neue Umgebung. Dabei wird z.B. ein MS Word 2010 Dokument in ein MS Word 2013 Dokument konvertiert. Bei der *Emulation* wird die ursprüngliche Umgebung der digitalen Objekte nachgestellt, d.h. hier passt sich die Umgebung den Daten an. Im Gegensatz dazu passen sich bei der *Migration* die Daten der Umgebung an.

Bei der *Migration* können *Datenmigration* (auch Formatmigration) und *Datenträgermigration* unterschieden werden. Letztere fand schon im Kontext der *Bitstream-Preservation* mehrfach Erwähnung und meint die Kopie von Daten von einem Trägermedium auf ein anderes, also z.B. von einer Festplatte auf einen USB-Stick. Bei der *Formatmigration* werden „Daten von einem Datenformat in ein aktuelleres, möglichst standardisiertes und offen gelegtes Format überführt“ (Funk 2010, 8.3). Dabei wird das digitale Objekt so modifiziert, dass seine Inhalte erhalten bleiben, es aber auf neuer Hardware benutzt werden

kann. Hierbei besteht allerdings immer das Risiko des Informationsverlusts, wenn z.B. das *Look-and-Feel* des Originals nicht mehr hergestellt werden können. Eine verlustfreie *Migration* ist nur bei genauer Kenntnis aller Spezifikationen des Original- und Zielformats (idealerweise Open Source) möglich und auch nur dann, wenn die Formate kompatibel sind. Mit wiederholter *Migration* in das jeweils neuere Format steigt jedoch die Wahrscheinlichkeit des Datenverlusts, weshalb immer auch die Originalversion mitgespeichert werden sollte, was aber einen erhöhten Speicherplatzbedarf mit sich bringt.

Bei der *Emulation* versucht man den bei der *Migration* möglichen Informationsverlust durch Nachahmung der originalen Umgebung der digitalen Objekte, zu verhindern. Es können Anwendungssoftware, Betriebssysteme und Hardware-Plattformen emuliert werden. Ein großer Vorteil der *Emulation* ist, dass die Originalobjekte unverändert erhalten bleiben und daher *Look-and-Feel* authentisch sind. Auch wird bedeutend weniger Speicherplatz wie bei einer *Migration* benötigt. Allerdings ist es oft sehr kosten- und zeitaufwendig, geeignete *Emulatoren* zu entwickeln, besonders, wenn nicht alle Details des ursprünglichen Objekts oder Systems offen zugänglich sind. Schließlich bleibt noch zu bedenken, dass dieser enorme Arbeitsaufwand für jede neue Generation von Hardware und Systemen neu betrieben werden müsste (Funke 2010, 8.4).

Das ursprünglich für das Gebiet der Raumfahrt von der Arbeitsgemeinschaft *Consultative Committee for Space Data System* (CCSDS) entwickelte Referenzmodell *Open Archival Information System (OAIS)*, welches auch als ISO-Standard anerkannt ist (ISO 14721:2003), verkörpert einen mittlerweile weltweit akzeptierten Standard zur Implementierung von Archivierungssystemen (Brübach 2010, 4.2) und kann die soeben beschriebenen Maßnahmen zur Langzeitarchivierung in ein offenes Rahmenmodell einfügen. Es ist ein logisches Modell, welches die technischen und organisatorischen Abläufe zur langfristigen Sicherung von Forschungsdaten vereinheitlicht und komplementär zu dem bereits vorgestellten *DCC Lifecycle Model* betrachtet werden kann: „Da es neutral gegenüber Datentypen und —formaten, Systemarchitekturen und Institutionstypen ist, bietet das OAIS-Modell eine wesentliche Grundlage, über Spartengrenzen hinweg über digitale Langzeitarchivierung und digitale Langzeitarchive sprechen zu können“ (nestor-Arbeitsgruppe OAIS-Übersetzung / Terminologie 2012, I).

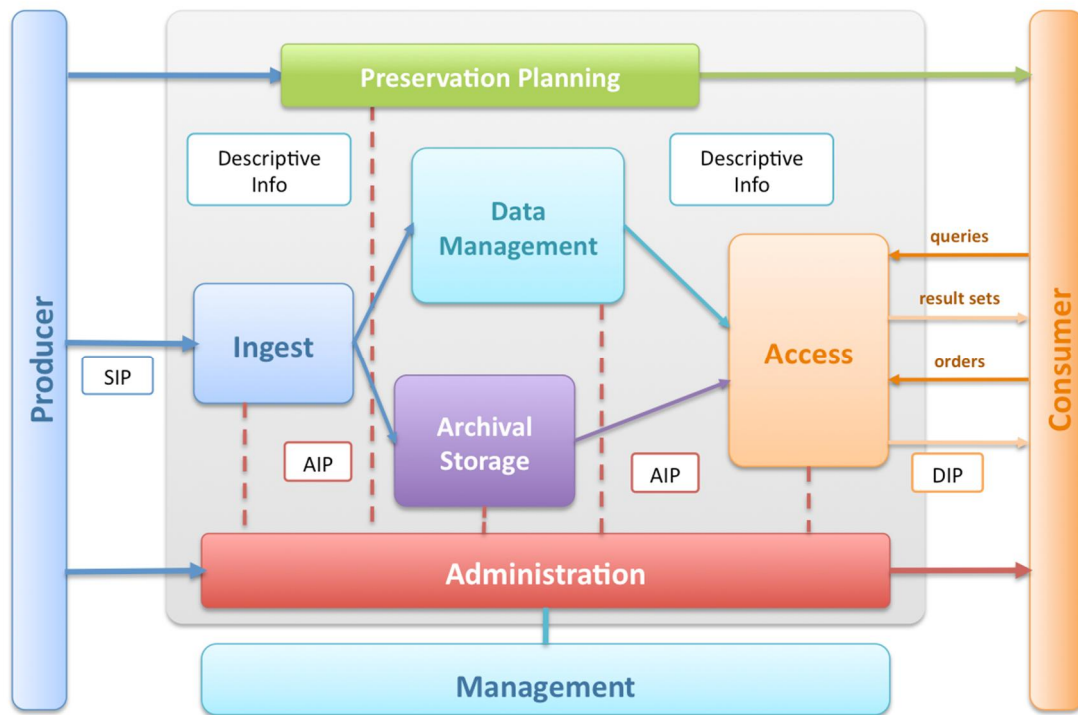


Abbildung 3: Das OAIS-Referenzmodell (Grafik von Fachlabor-Gubler.ch)

Das OAIS „beschreibt ein digitales Langzeitarchiv als eine Organisation, in dem Menschen und Systeme mit der Aufgabenstellung zusammenwirken, digitale Informationen dauerhaft über einen langen Zeitraum zu erhalten und einer definierten Nutzerschaft verfügbar zu machen“ (Brübach 2010, 4.2). Die Kernfunktionen eines Archivs sind hier die *Erhaltung* und der *Zugriff* auf die Daten. Die Erhaltung umfasst den Schutz der digitalen Objekte vor technischer Obsoleszenz und physischer Beschädigung (wie bereits erklärt) sowie die Sicherstellung der ausreichenden Beschreibung der digitalen Objekte durch entsprechende Metadaten. Der Zugriff beschäftigt sich mit den Maßnahmen, die ergriffen werden müssen, damit die digitalen Daten den Nutzern zugänglich gemacht werden können. Das Modell veranschaulicht, wie ein von einem Produzenten (*Producer*) erstelltes digitales Objekt (*SIP= Submission Information Package*) in das Archivierungssystem übernommen wird (*Ingest*). Nach der erfolgreichen Aufnahme in das System wird es nun durch Anreicherung mit Metadaten in ein Archivierungsobjekt (*AIP= Archival Information Package*) transformiert und im Archivspeicher (*Archival Storage*) abgelegt. Die Verwaltung des Objekts findet in der Datenverwaltung (*Data Management*) statt. Hierbei handelt es sich z.B. um Datenbank-Updates zur Sicherung von Lesbarkeit, Verständlichkeit und Nutzbarkeit oder die Wartung und das Zugänglichhalten der Verzeichnungsinformationen. Das archivierte digitale Objekt

wird nach einer Anfrage über den Zugriff (*Access*) an den Endnutzer (*Consumer*) unter Berücksichtigung eventueller (rechtlicher) Einschränkungen als Benutzerobjekt (*DIP= Dissemination Information Package*) ausgegeben. Die Langzeitarchivierungsstrategien (Migration und Emulation) werden in der Erhaltungsplanung (*Preservation Planning*) festgelegt. Die Verwaltung des kompletten Systems erfolgt in der Administration (*Administration*). Hierzu gehören unter anderem die Zuständigkeitsregelungen für die Arbeitsvorgänge im Archiv, die Qualitätssicherung und die fachmännische Pflege der Hardware.

Es wird immer deutlicher wie komplex die sorgfältige Langzeitarchivierung ist und so ist es auch nicht verwunderlich, dass sich derzeit viele Einrichtungen mit diesem Thema intensiv auseinandersetzen. Die Schwerpunktinitiative „Digitale Information“ beschäftigt sich im Rahmen des Handlungsfeldes „Nationale Hosting-Strategie“ noch mindestens bis 2017 mit der Langzeitarchivierung von digitalen Verlagspublikationen (Allianz der deutschen Wissenschaftsorganisationen: "Nationale Hosting-Strategie"). Der Wissenschaftsrat stellt 2011 in seinen *Empfehlungen zu Forschungsinfrastrukturen in den Geistes- und Sozialwissenschaften* fest, dass die WissenschaftlerInnen zwar gerne Daten von anderen ForscherInnen übernehmen, aber ungern selbst ihre Forschungsdaten archivieren und veröffentlichen und führt dies unter anderem auf die noch zahlreichen Probleme und ungelösten Fragen im Hinblick auf Langzeitarchivierung im Bereich der Geisteswissenschaften zurück (Empfehlungen zu Forschungsinfrastrukturen in den Geistes- und Sozialwissenschaften 2011, 54-57) und stellt fest: „Das Problem der Langzeitarchivierung digitaler wissenschaftlicher Informationen (Texte, Bilder, Film- und Tonaufnahmen, Forschungsprimärdaten, Quellcodes für Software) ist bislang auf nationaler Ebene nicht umfassend gelöst“ (ibid., 57). Daher empfiehlt der Wissenschaftsrat nachdrücklich „den Trägereinrichtungen die umfassende und langfristige Archivierung qualitätsgesicherter und für die jeweilige wissenschaftliche Gemeinschaft langfristig relevanter Daten“ (ibid. 58). Weiterhin stellt er fest: „Leicht bedienbare Benutzeroberflächen und Portale würden die Bereitschaft von Wissenschaftlerinnen und Wissenschaftlern, Daten zu archivieren, zweifellos erhöhen“ (ibid., 58). Schließlich unterstützt der Wissenschaftsrat den „Aufbau von Forschungsdatenzentren direkt bei den Datenproduzenten“ (ibid., 58) und fordert die Forschungsförderer ausdrücklich auf, „Anreize zu schaffen“ (ibid., 58), um die Bereitschaft

zur Archivierung zu erhöhen. Diese Anreize könnten laut Wissenschaftsrat in einer Optimierung der Förderung von Langzeitarchivierung bestehen, d.h.

1. Referenz- und entsprechende Zitationsmöglichkeiten für Datensätze sollten ausgebaut werden (ibid., 58).
2. Es sollten mögliche finanzielle Hindernisse bei der Datenarchivierung durch eine Förderung der anfallenden Personal- und Sachkosten ausgeräumt werden. Hier könnte z.B. die Deutsche Forschungsgemeinschaft (DFG) unterstützend auftreten (ibid., 59).
3. Die Nutzung von archivierten Daten könnte gefördert werden, indem die Durchführung von Re-, Sekundär- und Meta-Analysen förderfähig wird (ibid., 59).

Auch die *Kommission Zukunft der Informationsinfrastruktur* stellt 2011 in ihrem *Gesamtkonzept* zwar die besondere Bedeutung der *Data Curation* fest, beobachtet aber gleichzeitig noch großen Handlungsbedarf:

Derzeit ist die Situation noch geprägt durch mangelnde Transparenz über Verfügbarkeit und Qualität, fehlende Metadaten und Standardisierung. Das sogenannte Lebenszyklusmanagement ebenso wie der rechtliche und ethische Umgang mit Forschungsdaten stellen weitere Herausforderungen dar. (...) Datenerhaltung und Datenmanagement erfolgen nicht nach einheitlichen Kriterien oder übergreifenden Standards, und die Objekte sind bisher kaum in ausreichender Weise miteinander verknüpft. Ein **integriertes** und auf langfristige Verfügbarkeit ausgerichtetes **Informationsmanagement** ist somit ebenfalls eine große Herausforderung (Kommission Zukunft der Informationsinfrastruktur 2011, 17).

Die Kommission empfiehlt die Einbindung der Langzeitarchivierung direkt in die Virtuellen Forschungsumgebungen und Mehrwertdienste, Bereitstellung von Diensten zum vollständigen und dauerhaften Nachweis aller verfügbaren und relevanten Materialien (Metadaten, Persistent Identifier, Aggregationen, Kataloge), Festlegung verbindlicher Standards und Verfahren (Schnittstellen, Prozesse, Service Levels), Schaffung der notwendigen organisatorischen, technischen und finanziellen Rahmenbedingungen und stabile und eindeutige Verlinkungsmöglichkeiten zu ergänzenden Materialien (z.B. Forschungsdaten) (ibid., 33). Dabei definiert die Kommission Langzeitarchivierung als

„Sicherstellung des verlässlichen Zugriffs auf diese Inhalte „so lange wie möglich“, unabhängig von Speichermedienausfällen und (software-) technologischen Veränderungen“ (ibid., 33).

Die DFG empfahl bereits im Jahr 1998, dass Forschungsdaten für mindestens 10 Jahre archiviert werden sollten (DFG 1998, 12) und legte schließlich 2007 eine Aktionslinie „Entwicklung von Organisations- und Geschäftsmodellen“ auf. 2008 folgte KoLaWiss (Kooperative Langzeitarchivierung für Wissenschaftsstandorte), ein Projekt, welches die Anforderungen der wissenschaftlichen Community an digitale Langzeitarchivierungsmaßnahmen evaluieren sollte. Unter Projektleitung der Gesellschaft für wissenschaftliche Datenverarbeitung mbH (GWDG) „wurde untersucht, wie unter Einbeziehung von bereits vorhandenen Ansätzen verschiedene Akteure an einem Wissenschaftsstandort zum Aufbau einer gemeinsamen LZA-Struktur eingebunden werden können“ (KoLaWiss 2009). Man kam zu dem Ergebnis, dass eine kooperative Langzeitarchivierungsstruktur für Wissenschaftsstandorte nur dann realisiert werden kann, „wenn die entsprechenden technischen und organisatorischen Voraussetzungen bei den beteiligten Akteuren gegeben sind“ (ibid.). Das beinhaltet:

- Standardisierung und Nutzung von Metadaten für Forschungsdaten
- Policies und Workflows für LZA
- Software für LZA, insbesondere automatische Generierung von technischen Metadaten, Überwachung von Daten-Lebenszyklen, Schnittstellen zum Zugriff auf LZA-Daten
- Vereinbarungen über Service-Level zwischen LZA-Nutzern und LZA-Betreibern (...)
- Einrichtung von LZA-Büros bzw. LZA-Beauftragten an Wissenschaftsstandorten
- Zusammenarbeit und Arbeitsteilung zwischen LZA-Standorten
- Entwicklung von Zertifizierungskriterien und -verfahren für LZA-Knoten (ibid.).

Die DFG hat auch im Jahr 2009 die *Empfehlungen zur gesicherten Aufbewahrung und Bereitstellung digitaler Forschungsprimärdaten* veröffentlicht, die eine verantwortungsvolle Speicherung der Daten für mindestens 10 Jahre nun deutlich fordern.

Die für Deutschland wohl wichtigste Einrichtung mit Hinblick auf Langzeitarchivierung ist der ursprünglich vom Ministerium für Bildung und Forschung geförderte Kooperationsverbund *nestor -- Kompetenznetzwerk Langzeitarchivierung und Langzeitverfügbarkeit digitaler Ressourcen für Deutschland*. *nestor* vernetzt unterschiedliche Institutionen im Bereich LZA, kooperiert mit internationalen Partnern und erarbeitet mit seinen themenfokussierten Arbeitsgruppen Standards und Richtlinien, die auf seiner Informationsplattform veröffentlicht werden. Das Kompetenznetzwerk stellt in seinem Handbuch fest:

Es gilt also, jeweils geeignete Strategien für bestimmte digitale Sammlungen zu entwickeln, die je nach Bedarf und zukünftigem Nutzungsszenarium die langfristige Verfügbarkeit und Nachnutzung der digitalen Objekte sicherstellen. (...) „Langzeit“ bedeutet für die Bestandserhaltung digitaler Ressourcen nicht die Abgabe einer Garantieerklärung über fünf oder fünfzig Jahre, sondern die verantwortliche Entwicklung von Strategien, die den beständigen, vom Informationsmarkt verursachten Wandel bewältigen können (Liegmann 2010, 1).

2012 stellt *nestor* fest, dass „[für] den Bereich LZA von Forschungsdaten bislang keine nennenswerte Kooperationen [existieren]“ (Pempe 2012, 140). Weiterhin wird besonders mit Hinblick auf die Geisteswissenschaften und deren Forschungsinfrastrukturen konstatiert, dass es weder einheitliche Strategien noch übergreifende Lösungsansätze zur Langzeitspeicherung von Forschungsdaten gibt (ibid., 140). Den Auflagen der DFG zur Speicherung von 10 Jahren „kommen geisteswissenschaftliche Forschungsprojekte üblicherweise unter Rückgriff auf individuelle Lösungen nach (z.B. Ablage auf Projektrechner oder beim lokalen Rechenzentrum), wobei im günstigsten Fall allenfalls den Anforderungen der *Bitstream-Preservation* Rechnung getragen wird“ (ibid., 140). Positiv wird allerdings das Entstehen von Forschungszentren im Bereich *Digital Humanities* beobachtet (z.B. das Göttingen Centre for Digital Humanities), welche zukünftig Service und Support auf regionaler Ebene regeln könnten (ibid., 142). Die Heterogenität der Forschungsgegenstände und Forschungsdaten im Bereich *Digital Humanities* und die exponentiell wachsende Menge selbiger stellen ein bisher ungelöstes Problem dar.

The humanities and arts are the least likely of the disciplines to generate their own data in the forms of observations, models or experiments. Humanities scholars rely most heavily on

records, whether newspapers, photographs, letters, diaries, books, articles, records of birth, death marriage; records found in churches, courts, schools, and colleges; or maps. Any record of human experience can be a data source to a humanities scholar. (...) Data sources for humanities scholarship are growing in number and in variety, especially as more records are digitized and made available to the public (Borgman 2009, 33).

nestor ist der Auffassung, dass für die Geisteswissenschaften noch nicht einmal ein ungefährer Wert der langfristig zu speichernden Daten zu benennen ist, vermutet aber, dass ein Datenzentrum der Geisteswissenschaften sich sicherlich im Petabyte-Bereich bewegen würde (Pempe 2012, 148). „Die Eingrenzung auf einen „Kanon“ sowohl was die Inhalte, als auch was die Datentypen angeht, ist allenfalls ansatzweise möglich“ (ibid., 148).

Mit Hinblick auf all diese Hindernisse ist es nicht allzu verwunderlich, dass es noch nicht viele erfolgreiche Beispiele für Langzeitarchivierung innerhalb von Virtuellen Forschungsumgebungen im Bereich *Digital Humanities* gibt. Daher soll an dieser Stelle abschließend nur kurz ein Blick auf das bereits erwähnte *TextGrid Repository* geworfen werden, das dazu geeignet ist, zentrale Elemente zum Aufbau von Datenzentren beizutragen (ibid., 144).

TextGridRep Architektur

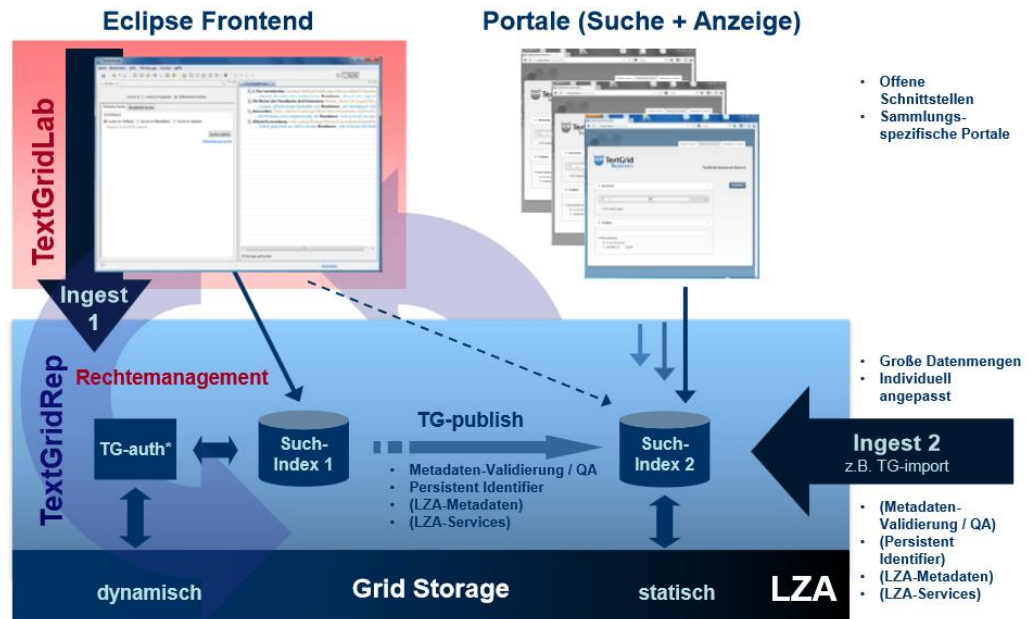


Abbildung 4: TextGridRep Architektur (Homepage TextGrid)

Es gibt im *TextGrid Repository* zwei separate Speicherbereiche mit zugehörigen Suchindizes. Im dynamischen Speicherbereich können die ForscherInnen ihre aktuell in Arbeit befindlichen Daten bearbeiten und über ein Rechtmanagement anderen Personen Rechte zum Lesen, Schreiben und Löschen zuweisen. Man kann dort immer nur die Daten suchen und sehen, für die man auch die Rechte besitzt. Der statische Bereich ist für publizierte, d.h. öffentliche Daten gedacht. Die Forschungsdaten können hier nicht mehr modifiziert werden und werden durch *Persistent Identifier* zitierbar. Diese Daten sind für jedermann über ein Internetportal suchbar. „Bei der Publikation von Forschungsdaten aus dem *TextGrid Laboratory* wird ein Prozess angestoßen, der automatisch Persistent Identifier vergibt, Format und Metadaten validiert, sowie die Daten im statischen, d.h. LZA-Bereich ablegt und dort nachweist“ (Pempe 2012, 145). Das *TextGrid Repository* strebt vier Service-Level an, die sich am *nestor Leitfaden für die Informationsübernahme in das digitale Langzeitarchiv* (nestor Materialien 10) orientieren (ibid., 146):

1. Einfache LZA (Bitstream-Preservation)	2. Einfache LZA in VFU	3. Content-Preservation	4. Data Curation
<ul style="list-style-type: none"> - Offline Speicherung - Redundante Speicherung - für 10 Jahre (DFG) 	Wie Level 1 + <ul style="list-style-type: none"> - Just-In-Time-Zugriff 	Wie Level 2 + <ul style="list-style-type: none"> - Technisches Qualitätsmanagement (Formatvalidierung, Formatkonvertierungen) 	Wie Level 3 + <ul style="list-style-type: none"> - kontinuierliche Pflege der Metadaten und der inhaltlichen Bezüge - Dissemination der Daten und Metadaten über Webschnittstelle

Tabelle 1: *TextGrid Repository* Service-Levels (nach Pempe 2012, 147)

Um diese verschiedenen Stufen umsetzen zu können, bedarf es aber noch der Klärung der rechtlichen Rahmenbedingungen und allgemeinen Nutzungsbedingungen.

2.5 Vergleichsstudien

Während der letzten fünf bis zehn Jahre gab es eine Vielzahl von Studien, die sich mit Forschungsinfrastrukturen und Langzeitarchivierung beschäftigt haben. Allerdings interessierte sich nur ein relativ geringer Teil davon dezidiert für die Geisteswissenschaften oder die speziellen Anforderungen dieser Fachdisziplin an die beiden Themen. Im Rahmen dieses Kapitels sollen nun einige besonders wichtige Studien kurz betrachtet werden, um den bisherigen Kenntnisstand darzustellen.

2004 wurde an der University of Rochester, NY (USA) eine relativ kleine Studie zur Erhebung der Bedürfnisse der dortigen ForscherInnen bezüglich ihrer Forschungs- und Schreibgewohnheiten durchgeführt, um die Publikationszahlen in ihren *Institutional Repositories* zu optimieren: „Our study focused on how faculty members do their research and writing. This entailed looking at how faculty members interact with digital tools and how they organize work in their virtual and physical workspaces“ (Foster und Gibbons 2005). Zu diesem Zweck wurden qualitative und quantitative Erhebungsmethoden kombiniert. Zwar waren dabei nicht nur geisteswissenschaftliche Forschungsdisziplinen

Untersuchungsgegenstand, jedoch sind die Ergebnisse richtungsweisend für die in dieser Arbeit diskutierten Entwicklungen bezüglich Forschungsinfrastrukturen. Foster und Gibbons stellen fest, dass eigentlich alle WissenschaftlerInnen, die sie befragten, digitale Werkzeuge in ihrem Arbeitsalltag benutzten, allerdings interessierte die meisten davon nur, „[that] these tools *did* work; they had little interest in *how* the tools worked or *what* the tools were“ (ibid.). Wie bereits dargestellt wurde, hat sich hier seit 2005 wohl nicht viel verändert. Weiterhin wurde als extrem wichtig empfunden, dass die Inhalte des *Institutional Repositories* langfristig auffindbar, nachnutzbar und zitierbar sein sollten. Gleichzeitig hatten die meisten Befragten keinen genauen Begriff von digitaler Datenverwaltung bzw. Langzeitarchivierung. Die meisten Probanden gingen davon aus, dass ihr Material bereits dauerhaft „sicher“ gespeichert sei. Metadatenstandards waren weitestgehend unbekannt und auch ein Bewusstsein für die digitale Formatierung ihrer Materialien war nicht vorhanden. Es ist zwar anzunehmen, dass sich die diesbezüglichen Kenntnisse der ForscherInnen in den letzten Jahren verbessert haben, doch wie bereits in den letzten Kapiteln festgestellt wurde, ist sich die Mehrheit der ForscherInnen der relevanten Fragen und Themen nicht hinreichend bewusst. Großes Interesse hatten alle Befragten daran, dass sie volle Kontrolle darüber behalten, wer Zugang zu ihren Materialien erhält – auch hier kein großer Unterschied zur gegenwärtigen Situation. Die beiden Autorinnen stellen fest, dass *Institutional Repositories* eigentlich nicht dazu geeignet sind, die Bedürfnisse der ForscherInnen zu unterstützen, nämlich den gesamten wissenschaftlichen Wertschöpfungsprozess zu fördern. So formulieren Foster und Gibbons in diesem Kontext schon die Qualitäten einer Virtuellen Forschungsumgebung:

These faculty members – and we believe them to be typical of faculty at most universities – are in desperate need of an authoring system to assist with document versioning, collaborative authoring, and centralized document access from any computer at any location. (...) It is with these [international] colleagues, many of them at other institutions, that researchers most want to communicate and share their work. But most organizations have mapped their IR communities to their departments rather than to the subtle, shifting communities of scholars engaged in interrelated projects (...) We believe that if we support the research process as a whole, and if faculty members find that the product meets their needs and fits their way of work, they will use it, and "naturally" put more of their work into the IR (ibid.).

Zwar wird hier noch von einem *Institutional Repository* gesprochen, allerdings dürfte ausreichend deutlich geworden sein, dass die Autorinnen eigentlich die Einrichtung eines modernen Forschungsinfrastruktursystems rund um den Forschungsprozess fordern.

Im Sommer 2004 führten Thomas Severiens und Prof. Dr. Eberhard Hilf am *Institute for Science Networking* in Oldenburg im Auftrag von *nestor* die Studie *Langzeitarchivierung von Rohdaten* durch. Methodisch wählten die Autoren eine Kombination aus qualitativen und quantitativen Methoden, d.h. einen Online-Fragebogen, vertiefende Interviews sowie einen Workshop. Es wurden 327 WissenschaftlerInnen und Institutionen, die zu großen Wissenschaftsorganisationen gehörten (z.B. Helmholtz-Gemeinschaft, Fraunhofer-Institute) oder im Kontext von Langzeitarchivierung von Forschungsprimärdaten damals aktiv waren, individuell angeschrieben. Die Studie kommt zu dem Schluss, dass obwohl eine kaum genau zu bestimmende, wachsende Zahl von Forschungsdaten erhoben wird, es wahrscheinlich aufgrund der Heterogenität der Fachdisziplinen, Datenformate, Metadaten und Schnittstellen keine einheitliche Möglichkeit geben kann, diese alle in eine übergreifende Langzeitarchivierungsinfrastruktur einbinden zu können (Severiens und Hilf 2004, 7). Bereits 2004 wurde festgestellt, dass es einer genauen Beschreibung der erhobenen Daten bedarf, also der „Dokumentation der Datenformate, [der] Methode der Datenerfassung und deren Dokumentation von Motivation und Zielsetzung“ (ibid., 7), damit diese überhaupt für andere WissenschaftlerInnen nachnutzbar sind. 63% der Befragten speicherten ihre Daten mit einer Selbstbeschreibung der Daten (ibid., 18). 45% der Probanden waren bereit, ihre Primärdaten anderen Kollegen außerhalb ihrer Institution zur Verfügung zu stellen, 28% stellten ihre Daten nur unter Vorbehalt zur Verfügung und weitere 27% lehnten dies schlicht ab (ibid., 18-19). Interessanterweise beantworteten nur wenige Institutionen die Frage nach vorhandenen Kooperationen mit externen Einrichtungen zur Langzeitarchivierung der eigenen Primärdaten, wobei leider keine genaue Zahl angegeben wird. Die Frage, ob es eine bestehende Kooperation zur Langzeitarchivierung der eigenen Daten gibt, wurde, wenn überhaupt, verneinend beantwortet; nur eine Institution bejahte dies (ibid., 19). Bei der Frage, ob rechtliche Rahmenbedingungen oder *DRM (Digital Rights Management)* als Hemmnis wahrgenommen werden, bejahten dies immerhin 45% der Befragten. Es wurde befürchtet, dass *DRM* den Zugriff auf die eigenen Daten verhindern könnte und eine allgemeine rechtliche Unsicherheit beklagt. Immerhin 75% bestätigten, schon einmal alte

(archivierte) Primärdaten genutzt und davon profitiert zu haben (ibid., 22). Schließlich forderten die Befragten Richtlinien und Hilfestellung zu folgenden Aspekten der Langzeitarchivierung: Datenformate, Dokumentationsrichtlinien, Nutzerschnittstellen, interne Datenverwaltung, *Persistent Identifier (PI)* und rechtliche Beratung (ibid., 23). Dies deckt sich auffallend mit den Handlungsfeldern, die auch heute noch, fast 10 Jahre später, von allen wichtigen Wissenschaftseinrichtungen identifiziert werden (wie bereits in den vorigen Kapiteln gezeigt), und noch immer nicht hinreichend gelöst wurden. Dies ist ein klarer Indikator dafür, wie komplex die Langzeitarchivierung wissenschaftlicher Daten ist.

2008 wurde an der *Hochschule für Technik und Wirtschaft Chur* unter Federführung von Prof. Hans-Dieter Zimmermann und Joachim Pfister eine von der SAGW (Schweizerische Akademie der Geistes- und Sozialwissenschaften) finanzierte *Bedarfsanalyse für ein Angebot „digitale Langzeitarchivierung“ in den Geisteswissenschaften (data repository)* durchgeführt. (Diese Studie war auch Vorbild für die im Rahmen dieser Masterarbeit durchgeführte Befragung. Einige Fragen wurden mit jeweiligen Abänderungen auch hier entlehnt.) Man bediente sich für diese Bedarfsanalyse der quantitativen Erhebungsmethode eines Online-Fragebogens. Befragt wurden 471 Personen in der Schweiz aus forschungsgetriebenen geisteswissenschaftlichen Infrastrukturen, alle geisteswissenschaftlichen universitären Institutionen/Institute sowie durch den SNF (der Schweizerische Nationalfonds) geförderte Projektleitende im Bereich Geisteswissenschaften. 89% der Befragten bejahten, dass digitale Daten in ihrer Einrichtung vorliegen (Zimmermann und Pfister 2008, 7), was die Notwendigkeit einer umfassenden Langzeitarchivierungsstrategie unterstreicht. Dies wird noch einmal deutlicher, da 20% der Befragten angaben, auf ältere digitale Objekte nicht mehr zugreifen zu können (ibid., 13). Dabei war mit 50% die häufigste Ursache, dass die Software zum Datenzugriff nicht mehr lief und bei 35% war die veraltete Hardware das Problem (ibid., 14). 98% der befragten ForscherInnen schätzen ihre Forschungsdaten als relevant für die Öffentlichkeit ein. Digitalisate sind bei 81% der Befragten vorhanden, digital erstellte (*born-digital*) Inhalte bei 73%, Datenbanken bei 70% und mit *OCR (Optical Character Recognition)* entstandene Daten bei 42% (ibid., 16). Die verwendeten Formate gliederten sich wie folgt auf: durchsuchbarer Text 82%, Bildformate 63%, Bilder, die Text beinhalten 62%, Audioformate 23% und Videodateien 14% (ibid., 16). Zum Erstellen und Bearbeiten der Dateien verwenden die WissenschaftlerInnen in der Regel (70%) Standardanwendungen wie

z.B. MS Office oder Open Office, wohingegen Web-Applikationen nur von 24% gewählt wurden (ibid., 28). Nur knapp die Hälfte aller Probanden (49%) erfasst Metadaten (ibid., 29), wobei nur ca. ein Drittel davon (34%) anerkannte Metadaten-Standards verwendet. 35% greifen auf nicht anerkannte Standards zurück und ein weiteres Drittel (31%) weiß nicht genau, wie es um seine Metadaten steht (ibid., 29). Ebenfalls sehr divers sind die Antworten bezüglich der Zuständigkeit für die Archivierung der Daten: 39% permanente Mitarbeiter, 32% Teilzeit-Beschäftigte, 33% Projektleitung, 26% Rechenzentrum, 12% externe Dienstleister (ibid., 32). Immerhin 92% der Befragten gaben an, dass Sicherungskopien der digitalen Objekte erstellt werden (ibid., 33). 41% geben an, dass die Daten nach Projektende zwecks Archivierung bei ihnen bleiben, 35% bleiben ohne Weiterbetreuung, 26% wissen noch nicht wie hier verfahren werden soll und 13% werden extern gesichert (ibid., 34). Nur 28% können bejahen, dass die Langzeitarchivierung der abgeschlossenen Projektdaten gesichert ist, wohingegen 50% dies verneint und 18% es schlicht nicht wissen (ibid., 35). Die Mehrheit wünschte Unterstützung bei der Langzeitarchivierung die sich aufgliederte in: 51% organisatorische Unterstützung, 58% technische Unterstützung, 50% Informationen über gängige Metadaten-Standards oder Initiativen innerhalb eines Fachgebiets und 58% finanzielle Unterstützung (ibid., 56). Die Frage nach der Bereitschaft zur Nutzung eines externen Dienstleisters zeigt ein gewisses Misstrauen der Befragten, da nur 13% dies ohne Einschränkung befürworten und 56% dies mit Einschränkungen erwägen würden (ibid., 58). Es bestätigt sich auch mit dieser Studie das bereits gezeichnete Bild einer sehr heterogenen Fachlandschaft, die sich der Komplexität der Langzeitarchivierung nicht ausreichend bewusst ist, nicht genügend Kenntnisse über Datenbeschreibung und internationale Standards hat und misstrauisch gegenüber Dritten ist, denen sie ihre digitalen Objekte anvertrauen sollen.

Zu dem gleichen Schluss kommt auch der Survey Report *Insight into Digital Preservation of Research Output in Europe (PARSE.Insight)*, der 2009 von Tom Kuipers und Jeffrey van der Hoeven veröffentlicht wurde. *PARSE.Insight (Permanent Access to the Records of Science in Europe)* war von 2008 bis 2010 ein EU-Projekt mit dem Ziel, eine Roadmap zur Unterstützung der e-Infrastruktur zur digitalen Langzeitarchivierung und Langzeitverfügbarkeit von wissenschaftlichen Primärdaten zu erarbeiten. Die Autoren des Reports stellen zusammenfassend fest:

From the researchers survey it is clear that researchers expect a significant growth of stored data. Likewise data managers too expect the amount of data they store at their facilities will grow significantly in the next five years. At the same time, however, the percentages of data managers who don't know what to expect in five years time grows to 28%. Perhaps this uncertainty can partly be explained by the current storage activities of researchers.

As became clear from the researchers survey, researchers tend not to store their digital data at external facilities. The most important storage locations for researchers were personal computers at work, portable storage carrier, organizational server, and computer at home. At the same time, as we have seen, researchers are concerned about legal issues and misuse of their data when it is stored elsewhere. One conclusion we can draw from this is that there is a human, psychological dimension to the issue of preservation which has to do with trust (Kuipers und van der Hoeven 2009, 76).

Es gilt also, nicht nur die technischen, organisatorischen, finanziellen und rechtlichen Rahmenbedingungen zu regeln, sondern auch, eine psychologische Komponente in Zusammenhang mit den Datenproduzenten zu berücksichtigen. So ist es, wie auch schon mehrfach erwähnt, unbedingt notwendig, die ForscherInnen von Beginn an in die Entwicklung von Virtuellen Forschungsumgebungen und die darin implementierten Langzeitarchivierungsstrategien miteinzubeziehen, damit die Scheu, diese auch aktiv zu nutzen, signifikant abnimmt.

Dieser Punkt wird auch von Dr. Annamaria Carusi und Dr. Torsten Reimer bei der Zusammenfassung der Ergebnisse ihrer *Virtual Environment Collaborative Landscape Study* von 2010 betont:

It is clear that the most effective way of approaching the development process of VREs is a **participatory mode of development**, with researchers closely involved in generating the requirements and evaluating their implementation. Development needs to occur in an iterative fashion, with constant feedback from researchers. There is also a need to support researchers through training opportunities (Carusi und Reimer 2010, 6).

Diese vom JISC (*Joint Information Systems Committee*) finanzierte Studie bediente sich einer Kombination von quantitativen und qualitativen Erhebungsmethoden in Form eines Online-Surveys, Interviews und Fallstudien. Das Online-Survey richtete sich an die weltweite

Forscher-Community, allerdings speisten sich die Ergebnisse hauptsächlich aus europäischen, nordamerikanischen und australischen Antworten (ibid., 9). Auch diese international angelegte Studie betont die Bedeutung der *Usability* der Angebote (ibid., 43). Daraus ergibt sich, die Virtuelle Forschungsumgebung und damit auch die darin implementierte Lösung zur Langzeitarchivierung müssen so konstruiert sein, dass sie möglichst einfach zu nutzen sind, da sonst die Hemmschwelle der WissenschaftlerInnen, sich damit auseinanderzusetzen, zu groß ist. Auch wird deutlich darauf hingewiesen, dass es unbedingt notwendig ist, die Wissenschaftscommunity aufzuklären, welche Vorteile sie durch Mitarbeit an und die aktive Arbeit mit einer Virtuellen Forschungsumgebung gewinnen können (ibid., 43). Es muss klar werden, dass die Daten langfristig besser in einer betreuten Forschungsinfrastruktur aufgehoben sind, als auf einem alten PC im Institut und gleichzeitig, dass dies nicht heißt, dass die WissenschaftlerInnen die Kontrolle über ihre Forschungsdaten und ihre Arbeit verlieren.

Zuletzt soll noch auf eine Befragung verwiesen werden, die Carol Tenopir mit einigen Kollegen in den USA 2011 unter dem Titel *Data Sharing by Scientists: Practices and Perceptions* veröffentlichte. Tenopir und Kollegen konzipierten einen Online-Fragebogen, den sie mit einer Schneeballmethode international (aber hauptsächlich in den Vereinigten Staaten) in den Wissenschaftseinrichtungen verteilten. Es gelang ihnen damit 1329 Antworten aus allen Forschungsdisziplinen zu erhalten, was die Ergebnisse sehr interessant macht. Auf die Frage, ob die ForscherInnen mit dem Prozess der Beschreibung ihrer Forschungsdaten zufrieden sind, bejahen dies 18% (Agree Strongly) bzw. 41,8% (Agree Somewhat) (Tenopir et al. 2011, 4), allerdings geben gleichzeitig 56,1% an, dass sie dabei keinem Metadatenstandard folgen (ibid., 6). Bezüglich des Vorhandenseins eines etablierten kurzfristigen Datenmanagementplans fallen die Antworten sehr verschieden aus: 17,2% Agree Strongly, 25,6% Agree Somewhat, 14,2% Neither Agree Nor Disagree, 20% Disagree Somewhat und 23,1% Disagree Strongly (ibid., 5). Langfristig sieht es da wenig besser aus: 15,6% Agree Strongly, 22,9% Agree Somewhat, 14,9% Neither Agree Nor Disagree, 21,1% Disagree Somewhat und 25,5% Disagree Strongly (ibid., 5). Es ist allerdings als sehr bedenklich zu erachten, dass ca. die Hälfte der Befragten angibt, dass die Daten langfristig wahrscheinlich nicht verantwortlich verwaltet und archiviert werden, was einen großen Verlust für die weltweite Wissenschaftliche Gemeinschaft bedeutet. Dies wird in dieser

Befragung auch durch Zahlen belegt. Immerhin 27% (Agree Strongly) bzw. 40% (Agree Somewhat) geben an, dass es für ihre Forschung nachteilig ist, nicht auf von Kollegen generierte Daten zugreifen zu können (ibid., 5). Auch mehr als die Hälfte der Probanden gaben an, dass es an ihrer Einrichtung kein Training zum Datenmanagement gib (26,5% Disagree Somewhat, 32,3% Disagree Strongly) (ibid., 5). Schließlich soll noch darauf hingewiesen werden, dass auch diese Studie bestätigt, dass die meisten ForscherInnen bereit wären, ihre Forschungsdaten zugänglich zu machen, sofern ein Rechtemanagement ihnen die Kontrolle über ihre Daten ließe (24% Agree Strongly, 39,6% Agree Somewhat) (ibid., 8).

2.6 Hypothesen

Sowohl die Lektüre der Vergleichsstudien als auch die Kenntnis der wissenschaftlichen Hintergrundinformationen führen zu folgenden Hypothesen für die im Rahmen dieser Arbeit durchgeführte Befragung:

1. Das Thema der Langzeitarchivierung betrifft alle WissenschaftlerInnen, da heute alle digitale Daten produzieren, sei es als MS Word-Dokument, als Bilddatei oder als eine Datei in anderen Formaten.
2. Die digitalen Forschungsdaten der WissenschaftlerInnen im Bereich *Digital Humanities* sind multimedial und liegen in heterogenen Dateiformaten vor.
3. In der Regel werden die Forschungsdaten nicht mit standardisierten Metadaten beschrieben.
4. Die WissenschaftlerInnen sichern ihre Daten entweder selbst oder gar nicht ab, d.h. sie vertrauen ihre Daten keinem Dienstleister an und es findet keine professionelle Langzeitarchivierung statt.
5. Die WissenschaftlerInnen haben Angst, die Kontrolle über ihre Daten zu verlieren.
6. Die WissenschaftlerInnen arbeiten bereits in individuellen („selbstgestrickten“) digitalen Forschungsinfrastrukturen miteinander.

3. Methodologie der Arbeit

Um sich einen Überblick darüber zu verschaffen, wie WissenschaftlerInnen der Digital Humanities derzeit Forschung betreiben, d.h. welche Datenformate sie tatsächlich benutzen, wie sie diese (langfristig) sichern, ob und wie sie die Datensätze kollaborativ beforschen und welche Anforderungen sie an eine Langzeitarchivierung dieser Daten stellen, bedürfte es einer umfassenden Befragung unter aktiven ForscherInnen dieser Wissenschaftsdisziplin an verschiedenen Standorten. Diese Befragung müsste sowohl qualitativ als auch quantitativ erfolgen, d.h. eine Kombination aus verschiedenen empirisch-sozialwissenschaftlichen Datenerhebungsmethoden, die sich gegenseitig ergänzen, also sogenannte 'Mixed-Methods Research'. „In general, mixed methods research represents research that involves collecting, analyzing, and interpreting quantitative and qualitative data in a single study or in a series of studies that investigate the same underlying phenomenon“ (Leech et al. 2009, 267).

Schon ziemlich lange hält sich hartnäckig das Vorurteil, dass nur quantitative Forschung (in Kontrast zu qualitativer Forschung) belastbare Ergebnisse produziert. Leech und Onwuegbuzie bemerken in 2009: „The quantitative research paradigm was considered 'research' because it was the first research paradigm that incorporated ontological, epistemological, axiological, rhetorical, and methodological assumptions and principles“ (Leech und Onwuegbuzie 2009, 266). Die quantitativen Erhebungsmethoden folgen einem Objektivitätsanspruch, der durch Einhaltung standardisierter Verfahren garantiert werden soll. Helmut Kromrey definiert die quantitative Forschung knapp als „ein streng zielorientiertes Vorgehen, das die ‚Objektivität‘ seiner Resultate durch möglichst weitgehende Standardisierung aller Teilschritte anstrebt und das zur Qualitätssicherung die intersubjektive Nachprüfbarkeit des gesamten Prozesses als zentrale Norm postuliert“ (ibid., 25). Standardisierung ist in diesem Kontext so zu verstehen, „dass die Erhebung der Information sich in jedem einzelnen Fall in gleicher Weise vollzieht“ (Kromrey 2009, 26).

Quantitative Methoden sind nützlich, um die Charakteristika der Welt, die wir untersuchen, zu dokumentieren und statistisch auszuwerten, um dann aufgestellte Hypothesen daran zu testen (Givens 2006, 378). Die gängigsten Methoden der Datenerhebung zur Messung zählbarer Eigenschaften sind standardisierte Interviews und Fragebogen, Beobachtung, und

Inhaltsanalyse. Die darauf basierenden Datensätze werden dann, entsprechend der anfänglichen Fragestellung, mit Hilfe diverser statistischer Methoden wie deskriptiver Statistik (z. B. Median, Varianz), Inferenzstatistik (beruhend auf Wahrscheinlichkeit z. B. Chi-Quadrat-Test), univariater Statistik (eindimensionale Häufigkeitsverteilung) und multivariater Statistik (z. B. Faktorenanalyse, Clusteranalyse) analysiert (Kromrey, 2009) und diese Ergebnisse dann mit der Haltbarkeit der ursprünglichen Hypothese abgeglichen.

Die qualitative Forschung stellt „dahingegen die Erkundung der Sinnsetzungs- und Sinndeutungsvorgänge der Akteure im Untersuchungsfeld, die Exploration kultureller Praktiken und Regeln und die genaue und tiefgehende Analyse und Beschreibung von Einzelfällen in den Mittelpunkt ihrer Forschungsbemühungen“ (Kelle 2008, 13). Das heißt hier ist der Fokus auf dem Individuum – die einzelne Person wird direkt (persönlich) befragt, wobei der persönliche, kulturelle und situative Kontext Einfluss finden.

Qualitative results are not designed to generalize or be replicable as with the results of quantitative studies. Rather, qualitative results often point to areas where single representation of reality does not hold, or where specific sub-populations have particular needs that are not reflected in broad generalizations. (Givens 2006, 380).

Der Einzelfall bildet den Bezugspunkt, wobei die gesamte Person den Untersuchungsgegenstand bildet. Es werden keine großen Fallzahlen benötigt, da schon Einzelfälle interessante Zusammenhänge aufzeigen können. Dies steht in Kontrast zur quantitativen Forschung, die möglichst große Fallzahlen untersucht, da sie danach strebt, generalisierende Aussagen machen zu können. Bei der quantitativen Forschung, welche sich um Repräsentativität bemüht, werden die Probanden oft durch eine Zufallsauswahl bestimmt. Die Auswahl muss auf jeden Fall vor der Durchführung der Untersuchung stattfinden, da aufgrund des Anspruchs auf Einheitlichkeit keine Veränderungen mehr am Sampling nach Beginn der Datenerhebung vorgenommen werden dürfen, anders bei qualitativen Studien, die i.d.R. keinen Repräsentativitätsanspruch haben. Hier werden die Probanden meist bewusst ausgewählt und interessante Zwischenergebnisse können dazu führen, dass das Sampling entsprechend angepasst wird. Die Untersuchung gleicht einem offen gestalteten Prozess, der auch die nachträgliche Anpassung der Fragestellungen ermöglicht. Ganz anders sollte dies bei quantitativen Studien sein, die darauf gründen, dass

der detaillierte Erhebungs- und Auswertungsplan bereits vor der Untersuchung vorliegen (Bernerburg 2006).

Die Methoden ergänzen sich dadurch, dass „quantitative tells us 'If'; qualitative tells us 'How or why' ” (Terrell 2012, 258). Sinnvoll erscheint im Kontext unserer Fragestellung eine Kombination aus einem persönlichen Interview, eines schriftlich auszufüllenden standardisierten Fragebogens und der Analyse der, der von den WissenschaftlerInnen tatsächlich archivierten Forschungsdaten, mit Hinblick auf Ihre langfristige Nutzbarkeit.

Im Rahmen dieser Masterarbeit kann jedoch nur eine Methode in vereinfachter Form Anwendung finden. Daher wurde die kosten- und zeitsparende quantitative Erhebungsmethode der Online-Befragung in Form eines Web-Surveys gewählt. Rainer Schnell stellt 2012 fest:

[...], [dass] seit einigen Jahren Befragungen über das Internet („Online-Befragungen“, „Web-Surveys“ etc.) zunehmend populär [werden]. Die Ursachen hierfür sind offensichtlich: Die Befragungen sind schneller durchführbar, man benötigt keine Interviewer, die erhobenen Daten müssen nicht erfasst werden, graphische Vorlagen hoher Komplexität sowie Audio- und Videosequenzen können im Erhebungsinstrument eingesetzt werden usw. Von besonderer Bedeutung sind die im Vergleich zu anderen Erhebungsmodi vernachlässigbaren Erhebungskosten (Schnell 2012, 290-291).

Durch den Einsatz einer web-basierten Umfrage kann in relativ kurzer Zeit, eine relativ große Anzahl von Personen aus der Zielgruppe befragt werden. Aufgrund der beruflichen Tätigkeit der Autorin an der Universitätsbibliothek Mainz liegt es nahe, ForscherInnen der Johannes Gutenberg-Universität Mainz für die Befragung heranzuziehen. (Das Sampling wird in Kapitel 4 näher beschrieben werden.) Dies erlaubt, falls nötig, direkten Kontakt für Rückfragen, schafft eine Vertrauensbasis auf Seiten der Befragten und ermöglicht ein zügigeres Sampling durch genauere Kenntnis der Informationsinfrastruktur der JGU. Da es sich um WissenschaftlerInnen handelt, ist davon auszugehen, dass sie über Internet verfügen und in der Lage sind einen solchen Fragebogen selbständig am PC zu beantworten. Natürlich ist davon auszugehen, dass auch bei optimaler Formulierung der Fragen und Durchführung der Umfrage nicht alle Probanden die Fragen gleich verstehen, was zur Verfälschung der Ergebnisse führen kann. Auch der persönliche Hintergrund der einzelnen Befragten und der

situative Kontext während des Ausfüllens (z.B. ein schreiendes Baby im Hintergrund) kann nicht mit in die Evaluierung der Ergebnisse einbezogen werden. Jedoch stellt die Gruppe, der Befragten eine relativ homogene Zielgruppe dar, sodass die Verzerrung (Bias), soweit im Rahmen dieser kleinen Befragung möglich, verhältnismäßig gering bleiben sollte. Alles in allem überwiegen doch für diese kurzfristige Studie die Vorteile eines Online-Fragebogens wie

...a lower likelihood of giving socially desirable responses [...] Further benefits of web-based administration include: elimination of scanning errors; elimination of multiple responses; having data immediately available in an electronic format even for free-text fields; and the possibility of including skip rules for irrelevant questions (Heiervang und Goodman 2011, 74).

Die so erhobenen Daten sollen im Anschluss mit Methoden der deskriptiven Statistik näher beleuchtet und in Zusammenhang mit den zu Beginn formulierten Hypothesen gebracht werden.

4. Methode der Befragung

Im Rahmen dieser Masterarbeit wurde zur Befragung der Zielgruppe ein Online-Fragebogen erstellt. Der Fragebogen wurde mit einer vom CMS für die HU angepassten und gepflegten Installation von LimeSurvey, einer Open Source Software zur Erstellung von Web-basierten Umfragen, erstellt. Die Umfrage und deren Ergebnisse werden in einer zentralen Datenbank des CMS (MySQL) abgelegt und stehen i.d.R. ein halbes Jahr zur Verfügung. Voraussetzung zur Nutzung ist ein HU-Account. Die Wahl dieses Umfrage-Tools mit einer HU-URL wurde bewusst gewählt, um das Misstrauen gegenüber der Umfrage zu minimieren.

Grundgesamtheit dieser personalisierten Befragung waren die WissenschaftlerInnen der Johannes Gutenberg-Universität Mainz des Fachbereichs 05 - Philosophie und Philologie (bestehend aus den Abteilungen: Department of English and Linguistics; Deutsches Institut; Institut für Allgemeine und Vergleichende Literaturwissenschaft; Institut für Buchwissenschaft; Institut für Film-, Theater- und empirische Kulturwissenschaft; Institut für Indologie; Institut für Slawistik; Philosophisches Seminar; Romanisches Seminar; Seminar für Orientkunde) sowie des Fachbereichs 07 – Geschichts- und Kulturwissenschaften (bestehend aus den Abteilungen: Vor- und Frühgeschichte; Ägyptologie und Altorientalistik; Klassische Archäologie; Klassische Philologie; Historisches Seminar; Kunstgeschichte; Musikwissenschaft; Ethnologie und Afrikastudien).

Für die Stichprobe ausgewählt wurden hiervon die Personen, die einen akademischen Grad über einem Dr. phil. erlangt haben, da anzunehmen ist, dass dieser Personenkreis bereits einige Forschung betrieben hat und auch weiterhin aktiv betreiben wird. Somit sollte also die nicht exakt überschaubare Zahl an Lehrbeauftragten und Doktoranden ausgeschlossen werden, die erst am Beginn ihrer wissenschaftlichen Forschungsarbeit stehen. Es wurden alle relevanten Mailadressen über das UnivIS (Informationssystem) der Universität Mainz bzw. die Homepages der Fachbereiche ermittelt. Dabei ergab sich für den Fachbereich 05 eine Gruppe von 98 Personen und für den Fachbereich 07 eine Gruppe von 75 Probanden. Insgesamt beinhaltete die Stichprobe somit 173 TeilnehmerInnen. Dies stellte die Gruppe von Probanden dar, die unter den gegebenen Umständen kurzfristig, aufgrund ihrer mehrjährig ausgeübten Tätigkeit, aussagekräftige Ergebnisse hervorbringen konnte.

When trying to decide whether the power of a test is adequate there are a couple of useful points to consider. Select the largest sample size you can sensibly test. If you have limited resources, time or access to subjects these restrictions may have priority. The best thing to do is to select the sample size you are able to test given these restraints, then test them. (Hinton 1995, 101)

Dieser ausgewählte Personenkreis erhielt eine Einladung via E-Mail mit dem Zugriffslink, die von einem Mitarbeiter-E-Mailaccount der Universitätsbibliothek Mainz mit entsprechender Signatur am 24.02.2013 verschickt wurde, um die Legitimität der Umfrage noch einmal deutlich zu machen. Die Probanden sollten wissen, dass die Ansprechpartnerin und Autorin der Befragung in ihrer direkten Umgebung zu finden ist und jederzeit kontaktiert werden kann. Der Mail voran ging jeweils ein Hinweis in den Fachbereichssitzungen der Fachbereiche am 16.01.2013 (FB 05) und 23.01.2013 (FB 07) die Befragung durch möglichst rege Teilnahme zu unterstützen. Am 10.02.2013 ging schließlich eine Erinnerungsmail an alle Probanden, um sie noch einmal aufzufordern an der Umfrage teilzunehmen, falls dies nicht schon geschehen sei, woraufhin noch einmal einige Antworten eingingen.

Da alle MitarbeiterInnen der Johannes Gutenberg-Universität Mainz ihren Uni-Mainz-Account für alle dienstlichen Belange nutzen müssen und über einen PC-Arbeitsplatz mit Internetzugang verfügen, war sichergestellt, dass die E-Mails auch die Zielgruppe erreichten und alle Zugang zur Umfrage erhalten konnten.

Die Befragung fand im Zeitraum vom 24.01.2013, 00:00 Uhr bis zum 24.02.2013, 24:00 Uhr statt. Dieser Zeitraum wurde bewusst gewählt, da davon auszugehen war, dass nun alle WissenschaftlerInnen nach den Feiertagen wieder an ihrem Arbeitsplatz waren und genug Zeit hatten, die Umfrage zu beantworten. Der Zeitraum von 4 Wochen wurde gewählt, um keinen Druck zu erzeugen und wirklich allen Personen die Möglichkeit zu geben, sich mit der Befragung auseinanderzusetzen bzw. Rückfragen stellen zu können.

Es wurde ein zur Verfügung gestelltes Style-Template mit dem Corporate Design der HU benutzt, welches die Seriosität der Umfrage noch einmal unterstrich, um auch besonders kritischen Personen in der Zielgruppe der angesprochenen WissenschaftlerInnen Rechnung zu tragen. Um die Umfrage vor SPAM-Attacken zu schützen wurde eine Sicherheitsfrage, in

Form einer einfachen Additionsaufgabe verwendet. Des Weiteren wurden Cookies für die Zugriffskontrolle benutzt, um sicher zu stellen, dass jeder Proband die Umfrage nur einmal beantwortete. Die Umfrage erfolgte vollständig anonym. Die Daten mit den Antworten enthielten keinerlei auf die Person zurückzuführende Informationen. Die TeilnehmerInnen konnten die teilweise fertiggestellte Umfrage jederzeit zwischenspeichern und zu einem späteren Zeitpunkt beenden. Die Befragten konnten an einem Fortschrittsbalken das Voranschreiten im Fragebogen verfolgen, während die Fragen in beliebiger Reihenfolge beantwortet werden konnten. Der Fragebogen wurde aufgrund der geringen Fragenanzahl nicht optisch in Themenblöcke unterteilt, wobei jedoch eine thematische Bündelung der Fragen erfolgte bzw. eine logische Abfolge eingehalten wurde.

Der Fragenbogen enthielt 17 bzw. 20 Fragen. 17 Fragen waren für alle TeilnehmerInnen zu beantworten, 3 Fragen wurden hinter Filter gelegt (Fragen 4, 12 und 13, siehe Logikdatei im Anhang), sodass sich je nach Antwort ggf. noch eine weitere Frage anschloss. Der komplette Fragebogen beinhaltete ausschließlich geschlossene Fragen und nahm ca. 10 Minuten in Anspruch. Es gab 3 Fragen mit einer Mehrfachauswahl und 17 Fragen mit Optionsfeldern. Hiervon wurden 13 Fragen mit einer 5-stufigen Likert-Skala (verbale Skalenbezeichnung: ja, eher ja, weder noch, eher nein, nein) zur „Messung der Einstellung“ (Likert-Skalierung, Gabler Wirtschaftslexikon) der Probanden gestaltet. Die Likert-Skala wurde zum besseren Verständnis noch einmal verbal kurz verdeutlicht (siehe Kopfzeile des Fragebogens). Die letzten 3 Fragen des Fragebogens erhoben demographische Angaben zu Geschlecht, Alter und akademischen Grad der TeilnehmerInnen.

Um sicherzustellen, dass alle Fragen verständlich sind und die Zeitangabe ausreichend ist wurde ein Pretest vorgeschaltet. Dabei gingen 4 Bibliothekarinnen der Universitätsbibliothek Mainz und 4 externe WissenschaftlerInnen aus den Geistes- und Geschichtswissenschaften den Fragebogen durch, wodurch gesichert wurde, dass der Wortlaut weitgehend für die Zielgruppe verständlich war. Anschließend wurde die Formulierung einiger Fragen modifiziert.

Die Analyse der Daten wurde mit dem integrierten Statistik-Tool von LimeSurvey und Microsoft Excel durchgeführt.

5. Ergebnisse der Befragung

5.1 Demographische Informationen

Insgesamt gab es 50 Antworten auf die Umfrage, d.h. 123 Personen haben nicht geantwortet. Die 50 Antworten unterteilen sich in 36 vollständige und 14 unvollständige Antworten. Die TeilnehmerInnen, die die Befragung nicht beendeten, haben ihre Zwischenergebnisse bis auf eine Ausnahme auch nicht zwischengespeichert, d.h. es liegen keine Antworten von diesen Personen vor. Daher werden sie im weiteren Verlauf dieser Auswertung auch nicht weiter berücksichtigt.

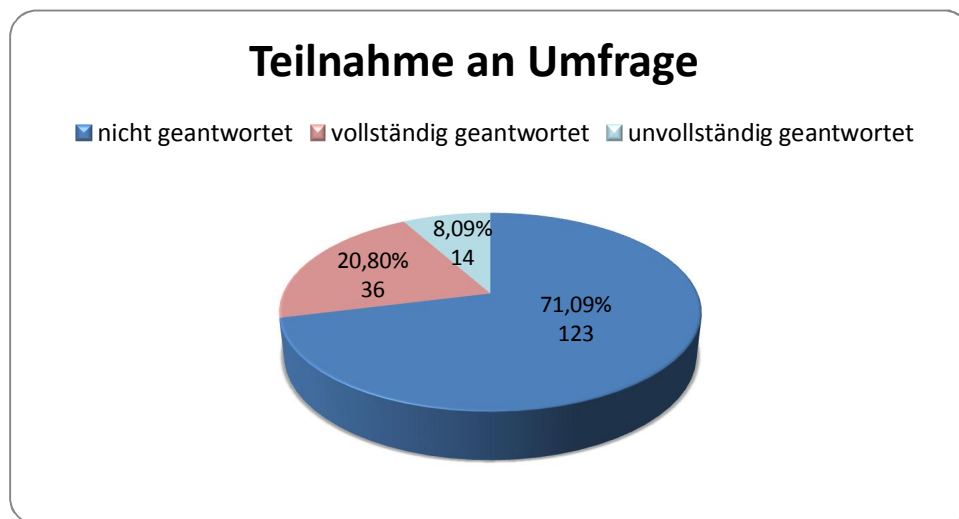


Abbildung 5: Teilnahme an Umfrage

Damit betrifft die Beendigungsquote dieser Umfrage insgesamt knapp 21%.

Innerhalb der angeschriebenen Zielgruppe waren 54 Personen weiblich und 119 Personen männlich. Von diesen 54 Frauen haben 13 vollständig geantwortet. Von den 119 Männern antworteten 23 vollständig. Damit liegt die Rücklaufquote bei den Frauen bei 24,07% und bei den Männern bei 19,32%. Da die Rücklaufquote bei männlichen wie weiblichen TeilnehmerInnen nicht weit auseinanderliegt, ist anzunehmen, dass das Interesse an dem Gebiet der Langzeitarchivierung auch annähernd gleich groß ist.

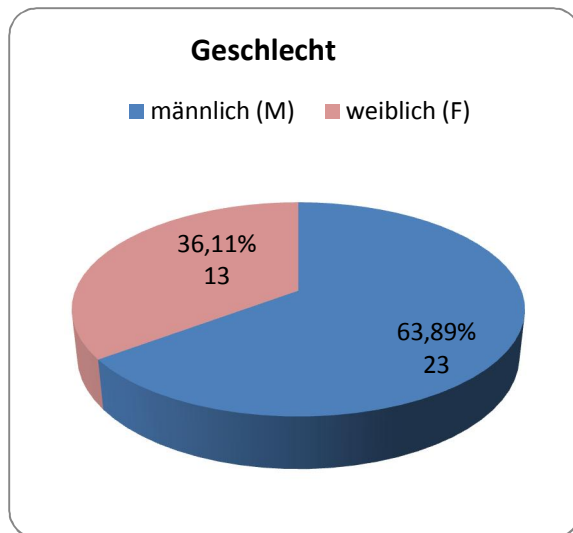


Abbildung 6: Geschlecht der TeilnehmerInnen

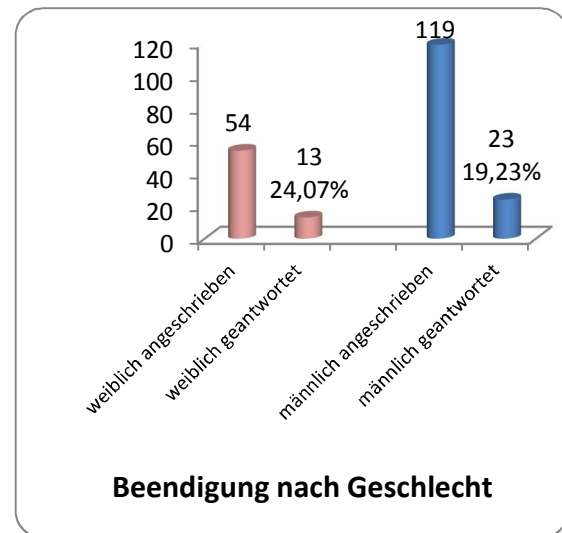


Abbildung 7: Beendigung nach Geschlecht

Bezüglich des akademischen Grads fällt auf, dass die meisten der TeilnehmerInnen, also insgesamt 41,66%, den Titel Prof. Dr. angegeben haben. Betrachtet man die Angaben geschlechterspezifisch, so entspricht dies bei den männlichen Befragten 52,17% und bei den weiblichen Befragten nur 23,07% aller befragten Wissenschaftlerinnen. Insgesamt kann man also sagen, dass die Mehrzahl der Befragten sehr erfahren in der Forschung ist. Sie können schon eine Habilitation vorweisen und kennen sich somit auch mit dem Forschungsprozess und der Forschungsdatenverwaltung aus. Erstaunlich ist die Tatsache, dass nur jeweils zwei Personen den Titel Juniorprofessor angaben, was die Vermutung nahelegt, dass die JungwissenschaftlerInnen sich noch nicht sehr intensiv mit den Fragen der Langzeitarchivierung beschäftigt haben bzw. noch nicht sehr viel Erfahrung mit der Problematik der Forschungsdatenverwaltung und –sicherung haben. Dies lässt sich darauf zurückführen, dass sie noch nicht sehr lange im Wissenschaftsbetrieb tätig sind und sie wahrscheinlich noch nicht oft direkt mit Datenverlust bzw. Datenunzugänglichkeit konfrontiert wurden.

Akademischer Titel der TeilnehmerInnen						
Titel	Anzahl ♀	Prozent ♀	Anzahl ♂	Prozent ♂	Anzahl total	Prozent total
Dr. habil.	0	0%	0	0%	0	0%
Juniorprof.	2	15,38%	2	8,68%	4	11,11%
PD Dr.	0	0%	0	0%	0	0%
PD Dr. habil.	2	15,38%	5	21,73%	7	19,44%
apl. Prof. Dr.	2	15,38%	4	17,39%	6	16,66%
Prof. Dr.	3	23,07%	12	52,17%	15	41,66%
Sonstige	4	30,76%	0	0%	4	11,11%
Summe	13	≈100%	23	≈100%	36	≈100%

Tabelle 2: Akademischer Titel der TeilnehmerInnen nach Geschlecht verteilt

Die Mehrzahl der Probanden ist also schon länger in der Wissenschaft aktiv tätig, was sich auch an der Altersverteilung ablesen lässt.

Altersverteilung						
Alter	Anzahl ♀	Prozent ♀	Anzahl ♂	Prozent ♂	Anzahl total	Prozent total
20 - 29	3	23,07%	0	0%	3	8,33%
30 - 39	3	23,07%	2	8,69%	5	13,89%
40 - 49	2	15,38%	9	39,13%	11	30,56%
50 - 59	5	38,46%	7	30,43%	12	33,33%
60 - 69	0	0%	5	21,73%	5	13,89%
Summe	13	≈100%	23	≈100%	36	≈100%

Tabelle 3: Altersverteilung nach Geschlecht

Das Durchschnittsalter liegt insgesamt bei 48 Jahren, wobei die weiblichen Teilnehmerinnen mit durchschnittlich 42 Jahren etwas jünger sind. Das Durchschnittsalter für die männlichen Teilnehmer liegt bei 52 Jahren.

Die Tatsache, dass hauptsächlich ältere WissenschaftlerInnen Interesse für das Thema dieser Umfrage zeigten, könnte auch damit in Zusammenhang stehen, dass viele von ihnen für große Forschungsprojekte verantwortlich sind und sich daher mit dem Thema der Forschungsdatenverwaltung direkt auseinandersetzen müssen. Wie bereits erwähnt, ist es derzeit bei Projektanträgen verpflichtend, Forschungsdatenmanagementpläne direkt miteinzureichen.

Erfreulich ist aber festzustellen, dass das Interesse für das Thema grundsätzlich in allen Altersgruppen vertreten ist, was deutlich macht, dass es sich hierbei um ein Thema handelt, dass den WissenschaftlerInnen wichtig ist.

5.2 Art der digitalen Objekte

Die Mehrheit der TeilnehmerInnen bejaht die Einstiegsfrage, ob sie digitale Forschungsmaterialien bearbeiten mit 91,67% (Modalwert), was 33 Personen entspricht. Nur 3 ForscherInnen (d.h. 8,33%) verneinen dies, was aber angesichts der Tatsache, dass alle WissenschaftlerInnen heute ihre Publikationen am PC verfassen, eigentlich nicht möglich ist. Festzuhalten ist, dass über 90% der Befragten definitiv mit digitalen Materialien arbeiten, d.h. es ist ein akuter Bedarf für das Forschungsdatenmanagement und deren Langzeitarchivierung vorhanden, womit Hypothese 1 eindeutig bestätigt wird.

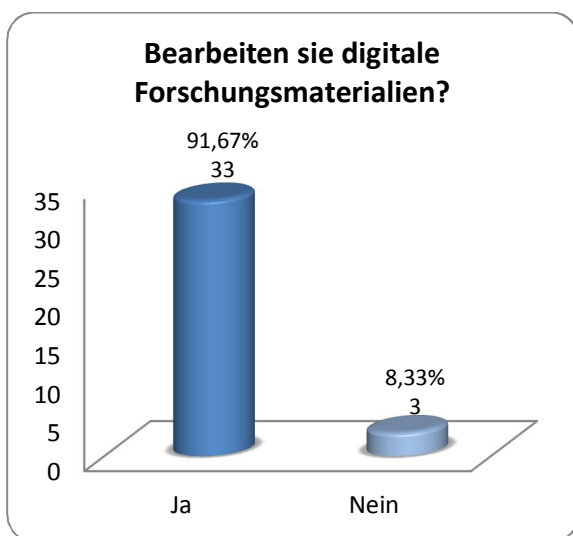


Abbildung 8: Bearbeitung von digitalen Forschungsmaterialien

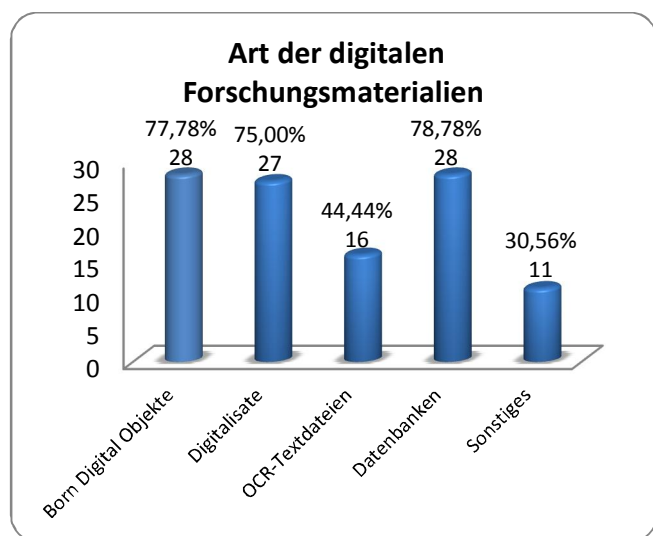


Abbildung 9: Art der digitalen Materialien
(Mehrfachnennungen möglich)

Mehrheitlich handelt es sich bei den digitalen Forschungsmaterialien um Materialien digitalen Ursprungs (*born digital*) 77,78% (d.h. 28 Nennungen), Datenbanken 78,78% (d.h. 28 Nennungen) sowie Digitalisate 75,00% (d.h. 27 Nennungen). Allerdings machen von mit OCR bearbeiteten Textdateien immerhin auch noch 44,44% (d.h. 16 Nennungen) Gebrauch.

Welche Dateiformate bearbeiten Sie? (Mehrfachnennung möglich)	Anzahl	Prozent
Digitale Bilder von Text (z.B. digitalisierte Schriftstücke, Bücher, ...)	28	77,78%
Digitale Bilder (z.B. von Fotos, Gemälden, Zeichnungen, Graphiken, ...)	23	63,89%
Durchsuchbare Textdateien (z.B. Word, PDF, XML, LaTeX, ...)	29	80,56%
Tondateien (z.B. wav, aup, mp3, Ogg Vorbis, ...)	11	30,56%
Videodateien (z.B. wmv, mpeg I u. II, mp4, avi, mov, flv, swf, ...)	13	36,11%
Sonstiges	8	22,22%

Tabelle 4: Frage zu Dateiformaten der Forschungsmaterialien

Besonders oft bearbeiten die Wissenschaftlerinnen durchsuchbare Textdateien (d.h. 29 Nennungen, also 80,56%), was auch zu erwarten war, da hierzu gängige Formate wie MS Word, PDF oder LaTeX gehören, die die ForscherInnen der Geisteswissenschaften zum Erstellen ihrer Publikationen benötigen. Gefolgt werden diese von digitalen Bildern von Text (Digitalisaten) mit 77,78% (d.h. 28 Nennungen) und digitalen Bildern von z.B. Fotos mit 63,89% (d.h. 23 Nennungen). Mit einigem Abstand findet man noch bei je grob einem Drittel Videodateien (d.h. 13 Nennungen, also 36,11%) und Tondateien (d.h. 11 Nennungen, also 30,56%). Wie sich feststellen lässt sind die Dateiformate sehr unterschiedlich. Die GeisteswissenschaftlerInnen arbeiten mit einer Bandbreite von multimedialen Objekten, die einer gezielten Pflege bedürfen, damit sie auch in einigen Jahren noch zugänglich und nachnutzbar sind.

5.3 Metadaten

Nur etwa ein Viertel der Befragten konnte positiv bestätigen, dass die Forschungsdateien mit Metadaten ausgezeichnet werden (*Ja*: 16, 67% bzw. 6 Nennungen und *Eher ja*: 13, 89% bzw. 5 Nennungen). Über 60% der Mainzer WissenschaftlerInnen benutzen keine Metadaten zur Beschreibung ihrer Daten (*Nein*: 38,89% bzw. 14 Nennungen und *Eher nein*: 27,78% bzw. 10 Nennungen). Dieser Wert übertrifft sogar die Resultate aus den Vergleichsstudien, in welchen etwa die Hälfte aller Probanden angab, keine Metadaten zu vergeben (vgl. Kapitel 2.5. Vergleichsstudien).

Frage	Ja		Eher ja		Weder noch		Eher nein		Nein		Keine Antwort	
	Anzahl	Prozent	Anzahl	Prozent	Anzahl	Prozent	Anzahl	Prozent	Anzahl	Prozent	Anzahl	Prozent
Verzeichnen Sie Metadaten zu ihren Forschungsdateien?	6	16,67%	5	13,89%	1	2,78%	10	27,78%	14	38,89%	0	0%
Sind dies nach anerkannten Standards vergebene Metadaten? (Frage hinter Filter)	2	5,56%	2	5,56%	1	2,78%	3	8,33%	3	8,33%	25	69,44%
Denken Sie, dass es für Ihre Arbeit wichtig ist Metadaten zu verzeichnen?	7	19,44%	9	25,00%	6	16,67%	8	22,22%	6	16,67%	0	0%

Tabelle 5: Fragen zu Metadaten

Von den WissenschaftlerInnen, die Metadaten vergeben, hält sich nur die Minderheit an standardisierte Metadaten (*Ja*: 5,56%, d.h. 2 Nennungen und *Eher Ja*: 5,56%, d.h. 2 Nennungen), was, wie schon in den vorherigen Kapiteln dargelegt, sehr problematisch für die Archivierung und Nachnutzbarmachung der Daten ist. Interessanterweise erachten es aber mit 19,44% *Ja*-Stimmen (d.h. 7 Nennungen) und 25% *Eher-ja*-Stimmen (d.h. 9 Nennungen) viele Befragte als wichtig, für ihre Arbeit Metadaten zu verzeichnen – ein Paradoxon. Was hält die WissenschaftlerInnen davon ab, die Metadaten zu vergeben, wenn sie so wichtig sind? Zum einen könnte da der Zeitfaktor eine Rolle spielen, da die ForscherInnen heute schon ein ungleich höheres Zeitpotential für administrative Tätigkeiten aufwenden müssen als es ihre Forschungstätigkeit erlaubt. Zum anderen herrscht, wie auch die Umfrageergebnisse zeigen, eine große Verwirrung bezüglich der Bedeutung von Metadaten und deren Standards, da immerhin 6 Personen keine eindeutige Antwort auf die Frage gaben (d.h. *Weder noch*: 16,67%), 8 mit *Eher nein* (d.h. 22,22%) und 6 mit *Nein* (d.h. 16,67%). Hier könnten Fortbildungsangebote, die sich an die ForscherInnen richten, helfen, die aufklären warum Metadaten so wichtig sind, welche Standards es gibt und wie sie sich auf ihre Daten anwenden lassen, und wie man Metadaten vergeben kann, ohne einen massiven Arbeitsaufwand zu generieren.

Es lässt sich also festhalten, dass sich auch die dritte Hypothese, nämlich dass die Forschungsdaten in den *Digital Humanities* in der Regel nicht mit standardisierten Metadaten beschrieben werden, in Mainz bestätigt.

5.4 Langzeitarchivierung bzw.-verfügbarkeit der digitalen Objekte

Die Mehrheit der ForscherInnen sprach sich mit 60% (d.h. 22 Nennungen) dafür aus, dass sie ihre digitalen Forschungsdaten definitiv in einigen Jahren noch benötigen werden. 22,22% (d.h. 8 Nennungen) sind sich ziemlich sicher und nur jeweils 3 Personen antworteten hier mit *Eher nein* und *Nein* (d.h. jeweils 8,33%). Dies legt nahe, dass, wenn die Daten langfristig benötigt werden, sie auch so gepflegt werden müssen, damit dieser Zugriff auch gewährleistet werden kann.

Frage	Ja		Eher ja		Weder noch		Eher nein		Nein	
	Anzahl	Prozent	Anzahl	Prozent	Anzahl	Prozent	Anzahl	Prozent	Anzahl	Prozent
Denken Sie, dass Ihre digitalen Dateien längerfristig, d.h. in einigen Jahren noch benötigen werden?	22	61,11%	8	22,22	0	0%	3	8,33%	3	8,33%
Ist bei Ihnen schon das Problem aufgetreten, dass Sie nicht mehr auf digitale Forschungsmaterialien zugreifen konnten, die Sie noch brauchten?	9	25%	4	11,11%	0	0%	10	27,78%	13	36,11%
Erfolgt derzeit eine Sicherheitskopie (Backup) Ihrer digitalen Daten?	16	44,44%	13	36,11%	1	2,78%	2	5,56%	4	11,11%
Würden Sie sich wünschen, dass Ihre Forschungsdaten auf unbestimmte Zeit sicher von Ihrer Universität archiviert werden?	22	61,11%	5	13,89%	1	2,78%	3	8,33%	5	13,89%
Würden Sie Ihre Daten einer wiss. Institution, die als Dienstleister auftritt, zum Speichern anvertrauen?	7	19,44%	9	25%	2	5,56%	10	27,78%	8	22,22%

Tabelle 6: Fragen zu Langzeitarchivierung

25% (d.h. 9 *Ja*-Nennungen) bzw. 11,11% (d.h. 4 *Eher ja*-Nennungen) der Befragten haben bereits nicht mehr auf Daten zugreifen können, d.h. hier wären dringend professionelle Langzeitarchivierungsmaßnahmen nötig gewesen, um den langfristigen Zugriff sicherzustellen. Die Mehrheit der WissenschaftlerInnen hat bisher noch keine Probleme mit dem Zugriff auf ihre Daten gehabt, doch ist bei den rasanten und ungewissen Entwicklungen im Bereich der digitalen Medien nicht sicher wie lange dies noch ohne *Data Curation* der Fall sein wird.

80% der befragten ForscherInnen (*Ja*: 44,44% mit 16 Nennungen und *Eher ja*: 36,11% mit 13 Nennungen) sichern ihre Daten mit einem Backup. Nur 6 Personen sichern ihre Daten nicht ab (*Nein*: 11,11%, d.h. 4 Nennungen und *Eher Nein*: 5,56%, d.h. 2 Nennungen) bzw. 1 Person (d.h. 2,78%) hat auf diese Frage nicht eindeutig geantwortet. Der Wunsch der Mehrheit mit 75% (*Ja*: 61,11%, d.h. 22 Nennungen und *Eher ja*: 13,89%, d.h. 5 Nennungen) ist, dass die Universität Mainz sich der langfristigen Archivierung der Forschungsdaten annimmt. Eine wissenschaftliche Institution als Dienstleister zur Speicherung der Daten zu verpflichten

sehen die Probanden sehr unterschiedlich. 44% (*Ja*: 19,44%, d.h. 7 Nennungen und *Eher ja*: 25% d.h. 9 Nennungen) würden diese Dienstleistung in Betracht ziehen und 40% (*Nein*: 22,22%, d.h. 8 Nennungen und *Eher nein*: 27,78% mit 10 Nennungen) stehen dem eher ablehnend gegenüber. Daher wäre es sinnvoll, eine Lösung zur Pflege und langfristigen Archivierung der Daten direkt an der Universität Mainz anzubieten, da hierfür eine breite Akzeptanz vorhanden wäre und langfristig die individuelle Sicherung der Daten durch die WissenschaftlerInnen selbst keine Zukunft mehr hat. Es wird deutlich, dass die ForscherInnen gerne eine professionelle Betreuung und Hilfestellung zur Langzeitarchivierung annehmen würden, aber gleichzeitig auch die Kontrolle über ihre Daten behalten möchten, indem sie weiterhin direkt an der Universität verortet bleiben. Somit bestätigen sich auch Hypothese vier und fünf.

5.5 Data Sharing

Frage	Ja		Eher ja		Weder noch		Eher nein		Nein		Keine Antwort	
	Anzahl	Prozent	Anzahl	Prozent	Anzahl	Prozent	Anzahl	Prozent	Anzahl	Prozent	Anzahl	Prozent
Denken Sie, dass Dritte (z.B. andere Wissenschaftler) Ihre digitalen Forschungsdateien zukünftig als hilfreich für weitere Forschungsprojekte erachten könnten?	17	47,22%	13	36,11%	3	8,33%	0	0%	3	8,33%	0	0%
Würden Sie Ihre digitalen Forschungsdateien mit anderen Wissenschaftlern teilen?	9	25,00%	15	41,67%	5	13,89%	4	11,11%	3	8,33%	0	0%
Würden Sie Ihre Forschungsdateien mit anderen teilen, wenn ein Rechtemanagement den Zugriff regeln würde? (Filter)	0	0%	2	5,56%	1	2,78%	1	2,78%	3	8,33%	29	80,56%
Teilen Sie vielleicht bereits jetzt schon Ihre digitalen Forschungsdateien mit anderen Personen?	14	38,89%	8	22,22%	0	0%	5	13,89%	9	25,00%	0	0%
Würden Sie Ihre digitalen Daten mit anderen Wissenschaftlern gem. beforschen?	17	47,22%	11	30,56%	0	0%	3	8,33%	5	13,89%	0	0%

Tabelle 7: Fragen zum Teilen von Forschungsdaten

83% (*Ja*: 47,22%, d.h. 17 Nennungen und *Eher ja*: 36,11%, d.h. 13 Nennungen) der Mainzer ForscherInnen sind sich sicher, dass ihre Forschungsdaten hilfreich für andere WissenschaftlerInnen sein könnten. 57% (*Ja*: 25,00%, d.h. 9 Nennungen und *Eher ja*: 41,67%, d.h. 15 Nennungen) der Befragten wären auch bereit, ihre Daten anderen Forschungsprojekten zur Verfügung zu stellen, wobei sich hierbei die Mehrheit nicht klar für *Ja*, sondern für *Eher ja* entschied, was gewisse Einschränkungen und Vorbehalte indiziert. 13,89% (d.h. 5 Nennungen) haben sich neutral hierzu geäußert und immerhin nur 3 Personen (d.h. 8,33%) lehnen dies strikt ab, bzw. 4 Personen antworteten hierauf mit *Eher nein* (d.h. 11,11%). Auf die Nachfrage an die Personen, die ihre Daten nicht teilen wollten, wäre die Mehrheit auch dann nicht dazu bereit ihre Daten zu teilen (*Nein*: 8,33%, d.h. 3 Nennungen), wenn ein Rechtemanagement den Zugriff auf die Daten regeln würde. Ein Großteil der Mainzer ForscherInnen arbeiten bereits jetzt schon aktiv mit Fachkollegen zusammen, da 61% (*Ja*: 38,89%, d.h. 14 Nennungen und *Eher ja*: 22,22%, d.h. 8 Nennungen) der Befragten angaben, ihre Daten mit anderen Personen zu teilen.



Abbildung 10: Frage zur Art des Teilens von Forschungsdaten
(Frage hinter Filter / Mehrfachnennungen möglich)

Die Mehrheit entscheidet sich hierbei zum Austausch via E-Mail (58,33%, d.h. 21 Nennungen). Mit einigem Abstand folgen der Austausch über Speichermedien wie CD-ROM oder USB-Stick (38,89%, d.h. 14 Nennungen), über ein gemeinsames Netzlaufwerk (38,89%, d.h. 14 Nennungen), über einen webbasierten „gemeinsamen Raum“ wie z.B. ein Wiki, einen Blog oder eine Cloud (30,56%, d.h. 11 Nennungen) oder über ein Fachportal (22,22% d.h. 8

Nennungen). Die Mainzer ForscherInnen sprechen sich mit 77%er (Ja: 47,22%, d.h. 17 Nennungen und Eher Ja: 30,56%, d.h. 11 Nennungen) deutlicher Mehrheit dafür aus, dass sie gerne digitale Daten mit anderen WissenschaftlerInnen gemeinsam beforschen würden. Lediglich 5 Personen (d.h. 13,89%) lehnten dies komplett ab und 3 Personen (d.h. 8,33%) antworteten hier mit *Eher nein*.

Es wird deutlich, dass die WissenschaftlerInnen auch jetzt schon ihren Arbeitsalltag aktiv in individuellen digitalen Forschungsinfrastrukturen mit anderen Fachkollegen gestalten, wie auch schon in Hypothese 5 richtig vermutet wurde. Sie tauschen Daten aus und beforschen diese auch gemeinsam, doch bedienen sie sich hierfür „selbstgestrickter“ Lösungen und speichern ihre gemeinsam erarbeiteten Daten nicht professionell. Daher wäre es zweckmäßig, hier eine modulare virtuelle Forschungsinfrastruktur für die Mainzer GeisteswissenschaftlerInnen anzubieten, die es Ihnen erlauben würde, innerhalb eines Rahmens gemeinsam zu arbeiten, ihre Daten professionell zu verwalten und schließlich durch Unterstützung von Fachpersonal auch langfristig sicher zu archivieren und nachnutzbar zu halten. Eine modulare virtuelle Forschungsumgebung könnte an die unterschiedlichen Forschungsvorhaben angepasst werden und die individuellen Bedürfnisse der WissenschaftlerInnen berücksichtigen. Der Bedarf scheint definitiv vorhanden zu sein und es ist auch klar geworden, dass sicherlich die Mehrheit der Befragten ein von der Universität betreutes Angebot gerne in Anspruch nehmen würde, wenn es ihnen ihre Arbeit erleichtert, aber trotzdem die Kontrolle über ihre Forschungsdaten lässt.

6. Fazit und Ausblick

Wir leben in sehr spannenden Zeiten – ganz besonders mit Blick auf die Entwicklung von Forschungsinfrastrukturen. Wie wir gesehen haben, werden wir gerade mit den Konsequenzen des Paradigmenwechsels von Gutenbergs gedrucktem Buch zur digitalen Gesellschaft konfrontiert, und niemand weiß genau, wo die Reise langfristig eigentlich hingeht:

We all experience it: a rising tide of information, sweeping across our professions, our families, our globe. As a society we now produce more information in one year than all the words spoken by all human generations before us. We create it, transmit it, store it, receive it, consume it – and then, often, reprocess it to start the cycle all over again. It gives us power unprecedented in human history to understand and control our world. But, equally, it challenges our institutions, upsets our work habits and imposes unpredictable stresses upon our lives and societies (High Level Expert Group on Scientific Data 2010, 7)

Alles scheint im Fluss und so ist die Verunsicherung auch unter den WissenschaftlerInnen groß, wie sie mit den neuen digitalen Forschungsdaten und den daraus resultierenden Problemen für den Forschungsprozess umgehen können.

Die im Zusammenhang mit dieser Masterarbeit durchgeführte Befragung hat deutlich gezeigt, dass die geisteswissenschaftlichen ForscherInnen der Universität Mainz sich mehrheitlich auf digitale, multimediale Daten in heterogenen Dateiformaten einlassen, diese auch gemeinschaftlich mit Fachkollegen beforschen würden bzw. dies auch schon über individuelle Methoden und Kanäle tun. Daher wäre es zweckmäßig, den WissenschaftlerInnen einen Rahmen anzubieten, innerhalb dessen sie, durch Werkzeuge und Infrastruktur einer Virtuellen Forschungsumgebung unterstützt, ihre Forschungsfragen kollaborativ, orts- und zeitungebunden bearbeiten können. Durch eine solche Maßnahme könnten die WissenschaftlerInnen nachdrücklich in ihrem Arbeitsalltag entlastet werden. Dabei wäre es empfehlenswert, eine modulare Architektur zu wählen, die die Anpassung an verschiedenste Forschungsprojekte erlaubt. Die Forschungswerkzeuge müssen in enger Zusammenarbeit mit den WissenschaftlerInnen entwickelt werden, sodass deren *Usability* und Akzeptanz durch die Nutzenden sichergestellt werden können.

Die Mainzer WissenschaftlerInnen, wie auch ihre internationalen Kollegen und Kolleginnen in den Vergleichsstudien, schätzen die Relevanz ihrer Forschungsdaten für zukünftige Projekte hoch ein und sichern ihre Daten daher individuell ab. Allerdings werden die Daten nicht ausreichend mit standardisierten Metadaten beschrieben und auch nicht professionell langfristig gespeichert, da das Fachwissen über *Data Curation* sehr speziell, international noch nicht einheitlich definiert und unter Wissenschaftlern (wie gesehen) noch nicht sehr verbreitet ist. Es bietet sich an, diese professionelle Pflege und Verwaltung der Daten in Kooperation mit Fachpersonal zu gestalten, sogenannten *Data Curators*. Wie in dieser Arbeit hoffentlich deutlich wurde, ist die Pflege, Verwaltung und Archivierung wissenschaftlicher Forschungsdaten ein sehr komplexes und noch am Anfang stehendes Fachgebiet der Informationswissenschaft, da eine Vielzahl von rechtlichen, finanziellen und technischen Fragen noch zu beantworten sind. Es ist den WissenschaftlerInnen in der heutigen Zeit, die sich durch eine Datenflut charakterisiert, nicht zuzumuten, sich alleine in dem Urwald von Standards, Normen und Systemen zurechtzufinden. Vielmehr ist eine fundierte Kenntnis von informationstechnischem, rechtlichem als auch forschungsdisziplinspezifischem Fachwissen notwendig, um eine langfristige Archivierung und Nachnutzbarkeit eindeutig identifizierbarer Daten gewährleisten zu können. Es wird zu entscheiden sein, ob die Daten nur einer *Bitstream Preservation*, einer technischen Nachnutzbarkeit oder sogar einer inhaltlichen Nachnutzbarkeit zugeführt werden sollen. Die damit zusammenhängenden Maßnahmen müssen entsprechend organisiert werden und gegebenenfalls müssen die Daten mit ausreichend Hintergrund- und Kontextinformationen in standardisierte Form beschrieben werden, sodass eine inhaltliche Nutzung der Daten überhaupt erst möglich wird. Auch ist spezielles Fachpersonal, als Schnittstelle zur gemeinschaftlichen Entwicklung von Werkzeugen, für die Virtuellen Forschungsumgebungen unverzichtbar.

Die mehrheitliche Bereitschaft der Befragten, ein Langzeitarchivierungsangebot der Universität Mainz anzunehmen, legt nahe, dass die Forschenden durchaus Unterstützung bei der Sicherung ihrer Daten annehmen würden, solange garantiert bleibt, dass sie die Kontrolle über ihre Daten behalten und diese möglichst nicht Dritten anvertraut werden müssen. Die ForscherInnen haben ausgedrückt, dass sie ihre Forschungsdaten als wichtig für andere Forschungsprojekte erachten, und stehen daher prinzipiell der professionellen, langfristigen Archivierung ihrer Forschungsdaten positiv gegenüber. Auch hier entsprechen die Ergebnisse

der Befragung in Mainz in wesentlichen Teilen den Resultaten aus den internationalen Studien.

Die WissenschaftlerInnen benötigen also eine Lösung „aus einem Guss.“ So ist es auch auf europäischer Basis angestrebt die Langzeitarchivierung in Virtuelle Forschungsumgebungen zu integrieren, sodass der gesamte wissenschaftliche Wertschöpfungsprozess von einer Infrastruktur unterstützt werden kann – eine flexible One-Stop-Lösung. Eines der wenigen bereits existierenden Beispiele in nationalem Kontext hierfür ist das bereits beschriebene *TextGrid*, ein geisteswissenschaftliches Projekt in Zusammenhang mit dem vom BMBF (Bundesministerium für Bildung und Forschung) geförderten *WissGrid-Projekt*. Ganz aktuell veröffentlichte *WissGrid* Anfang 2013 einen *Leitfaden zum Forschungsdaten-Management*, da man die akute Notwendigkeit der Klärung allgemeiner Fragen erkannte: „Zwar gibt es im anglo-amerikanischen Bereich bereits eine Reihe von Planungsinstrumenten für das Forschungsdaten-Management, aber es gibt bisher keine vergleichbaren deutschsprachigen Hilfsmittel und aufgabenorientierte Einführungen“ (Ludwig und Enke 2013, 5).

Im Februar 2013 startete das EU-Projekt *PERICLES (Promoting and Enhancing reuse of Information throughout the Content Lifecycle taking account of Evolving Semantics)*, welches unter der Führung des King's College, London und unter Mitarbeit von sechs internationalen Partnern, wie z.B. der SUB Göttingen, sich verändernde Semantik der zu archivierenden digitalen Daten untersuchen wird: „PERICLES will be researching ways of modelling changing environments and of allowing software to capture and use contextual information to ensure that digital content remains accessible and understandable in the long term“ (King's College London 2013). *PERICLES* möchte einen *preservation-by-design*-Ansatz entwickeln, d.h. die notwendigen Abläufe zur Langzeitarchivierung sollten zukünftig in den Forschungsalltag integriert werden und nicht mehr wie bisher meist, a-posteriori stattfinden. Das bis 2017 geförderte EU-Projekt entstand aus dem Verständnis, dass die Langzeitarchivierung und die dazu notwendigen Prozesse der exponentiell anwachsenden Forschungsdatenflut zukünftig möglichst automatisiert ablaufen müssen, nicht zuletzt aufgrund der enormen finanziellen Herausforderung der Langzeitarchivierung.

In diesem Zusammenhang wurde ebenfalls im Februar 2013 ein weiteres EU-Projekt ins Leben gerufen, dass die Kosten des langfristigen Erhalts der Datenbestände fokussiert. 4C – 60

Collaboration to Clarify the Costs of Curation wird unter Mitarbeit der Deutschen Nationalbibliothek und zwölf weiterer europäischer Partnerinstitutionen im Zeitraum bis 2015 versuchen, Kostendaten und –modelle zu erarbeiten, die zukünftig die Langzeitarchivierung von Forschungsdaten an wissenschaftlichen Einrichtungen besser planbar und kalkulierbar machen sollen (4C "Overview" 2013).

Derzeit wird auf Bundesebene aktiv daran gearbeitet, wie in der im April 2013 veröffentlichten *Roadmap für Forschungsinfrastrukturen* des BMBF (Bundesministerium für Bildung und Forschung) dargestellt, wegweisende Forschungsinfrastrukturen zu fördern, die die „Forschungsgesellschaft für die nächsten 10-15 Jahre gestalte[n]“ (BMBF 2013, 2) werden. Für den Bereich der *Digital Humanities* gehören hierzu weiterhin das bereits vorgestellte *DARIAH-DE*, das sich von 2013 bis 2017 in der Aufbauphase befindet und ab 2018 in den regulären Betrieb übergehen soll, als auch *CLARIN-D* einer Infrastruktur, die sprachliche Forschungsfragen in den Vordergrund stellt. Es ist die Hauptaufgabe von *DARIAH-DE* „die Interoperabilität von Werkzeugen und Forschungsdaten sowie deren Langzeitverfügbarkeit und Nachnutzung in den *Digital Humanities* zu ermöglichen“ (*ibid.*, 17). *CLARIN-D* ist ein Zentrenverbund für WissenschaftlerInnen der Geistes- und Sozialwissenschaften. Zentrale Aufgabe von *Clarín-D-Zentren* ist es:

große Mengen relevanter Text- und multimedialer Sprachressourcen [zu pflegen] und über das Internet für Forschungszwecke [bereitzustellen] (lokal, national, europäisch); web-basierte Software-Werkzeuge und Technologien für disziplinen- und institutionenübergreifende Zusammenarbeit zur Verfügung [zu stellen], um die Masse an elektronisch verfügbaren Daten strukturieren und beherrschen zu können [sowie] de[n] Wandel der Sprachen und Kulturen durch die Archivierung sprachlicher Daten für zukünftige Generationen [zu erfassen] (*ibid.*, 16).

Die Aufbauphase hierfür reicht von 2012 bis 2017 und ab 2018 soll der reguläre Betrieb möglich sein.

Dies sind nur wenige Beispiele für die bereits existierenden Initiativen zur Gestaltung einer modernen und nachhaltigen Forschungsinfrastruktur für die nächsten 10-20 Jahre. Allerdings ist ungewiss, wie sich langfristig der wissenschaftliche Forschungsalltag entwickeln wird, da

ständig neue Entwicklungen, wie z.B. erste erfolgreiche Versuche zur Speicherung von Daten in DNA-Strängen (Goldman et al. 2013), andeuten.

7. Abbildungsverzeichnis

Abbildung 1: Der wissenschaftliche Wertschöpfungszyklus.....	13
Abbildung 2: Curation Lifecycle Model (DCC 2010)	21
Abbildung 3: Das OAIS-Referenzmodell (Grafik von Fachlabor-Gubler.ch)	26
Abbildung 4: <i>TextGridRep</i> Architektur (Homepage <i>TextGrid</i>)	32
Abbildung 5: Teilnahme an Umfrage	48
Abbildung 6: Geschlecht der TeilnehmerInnen.....	49
Abbildung 7: Beendigung nach Geschlecht	49
Abbildung 8: Bearbeitung von digitalen Forschungsmaterialien	51
Abbildung 9: Art der digitalen Materialien (Mehrfachnennungen möglich).....	51
Abbildung 10: Frage zur Art des Teilens von Forschungsdaten (Frage hinter Filter / Mehrfachnennungen möglich).....	56

8. Tabellenverzeichnis

Tabelle 1: <i>TextGrid Repository</i> Service-Levels (nach Pempe 2012, 147)	33
Tabelle 2: Akademischer Titel der TeilnehmerInnen nach Geschlecht verteilt.....	50
Tabelle 3: Altersverteilung nach Geschlecht	50
Tabelle 4: Frage zu Dateiformaten der Forschungsmaterialien	51
Tabelle 5: Fragen zu Metadaten	52
Tabelle 6: Fragen zu Langzeitarchivierung	54
Tabelle 7: Fragen zum Teilen von Forschungsdaten	54

9. Literaturverzeichnis

American Council of Learned Societies (ACLS). *Our cultural commonwealth*. Report of the ACLS Commission on Cyberinfrastructure for Humanities and Social Sciences. New York: ACLS, 2006.

<http://www.acls.org/cyberinfrastructure/ourculturalcommonwealth.pdf> (zuletzt abgerufen 11.11.2013).

Allianz der deutschen Wissenschaftsorganisationen. Schwerpunktinitiative „Digitale Information“, "Definition Virtuelle Forschungsumgebung."

http://www.allianzinitiative.de/de/handlungsfelder/virtuelle_forschungsumgebungen/definition/ (zuletzt abgerufen 11.11.2013).

Allianz der deutschen Wissenschaftsorganisationen. Schwerpunktinitiative „Digitale Information“, "Forschungsdaten."

<http://www.allianzinitiative.de/de/handlungsfelder/forschungsdaten/> (zuletzt abgerufen 11.11.2013).

Allianz der deutschen Wissenschaftsorganisationen. Schwerpunktinitiative „Digitale Information“, "Nationale Hosting-Strategie."

http://www.allianzinitiative.de/de/handlungsfelder/nationale_hosting_strategie/ (zuletzt abgerufen 11.11.2013).

Aschenbrenner, Andreas, Tobias Blanke, Stuart Dunn, Martina Kerzel, Andrea Rapp und Andrea Zielinski. "Von e-Science zu e-Humanities – Digital vernetzte Wissenschaft als neuer Arbeits- und Kreativbereich für Kunst und Kultur." *Bibliothek Forschung und Praxis*, Jahrgang 31, No. 1 (2007): 11-21.

http://www.b2i.de/fileadmin/dokumente/BFP_Bestand_2007/Jg_31-Nr_1/Jg_31-Nr_1_Aufsaeetze/Jg_31-2007-Nr_1-S_11-21.pdf (zuletzt abgerufen am 11.11.2013)

Bemerburg, Ivonne. *Methoden quantitativer Sozialforschung I –*

Datenerhebungsmethoden, (2006). <http://www.hitzler-soziologie.de/seminare/ws0607/Sozialforschung.htm> (zuletzt abgerufen 11.11.2013).

Berliner Erklärung über offenen Zugang zu wissenschaftlichem Wissen, 22.10.2003.

http://www.bideutschland.de/download/file/download/D5_SERVICE-Download_OpenAccess-Berliner-Erklaerung2003.pdf (zuletzt abgerufen 11.11.2013).

BMBF – Bundesministerium für Bildung und Forschung. "Roadmap für

Forschungsinfrastrukturen," 2013. <http://www.bmbf.de/pub/Roadmap.pdf> (zuletzt abgerufen 11.11.2013).

Borgman, Christine L. "The Digital Future is Now: A Call to Action for the Humanities." *Digital Humanities Quarterly*, Vol. 3, No. 4 (2009).
<http://www.digitalhumanities.org/dhq/vol/3/4/000077/000077.html> (zuletzt abgerufen 11.11.2013).

Brübach, Nils. "4.2 Das Referenzmodell OAIS" *nestor-Handbuch. Eine kleine Enzyklopädie der digitalen Langzeitarchivierung*. Version 2.3., herausgegeben von Heike Neuroth, Achim Oßwald, Regine Scheffel, Stefan Strathmann und Karsten Huth, 2010.
<http://nestor.sub.uni-goettingen.de/handbuch/nestor-handbuch.pdf> (zuletzt abgerufen 11.11.2013).

Büttner, Stephan, Hans-Christoph Hobohm und Lars Müller. "Research Data Management." *Handbuch Forschungsdatenmanagement*, herausgegeben von Stephan Büttner, Stephan, Hans-Christoph Hobohm und Lars Müller. Bad Honnef: Bock + Herchen Verlag, 2011: 13-24.

4C Project – the Collaboration to Clarify the Costs of Curation. "Overview," 2013.
<http://4cproject.net/overview/> (zuletzt abgerufen 11.11.2013).

Carusi, Annamaria und Torsten Reimer. "Virtual Research Environment Collaborative Landscape Study," 2010.
<http://www.jisc.ac.uk/media/documents/publications/vrelandscape.pdf> (zuletzt abgerufen 11.11.2013).

DARIAH-DE – "Projektziele." <http://de.dariah.eu/projektziele.html> (zuletzt abgerufen 13.05.2013).

DARIAH-EU - Digital Research Infrastructure for the Arts and Humanities. "Home."
http://www.dariah.eu/index.html?option=com_content&view=frontpage&Itemid=113 (zuletzt abgerufen 11.11.2013).

Digital Curation Centre – DCC. "DCC Curation Lifecycle Model," 2010.
<http://www.dcc.ac.uk/resources/curation-lifecycle-model> (zuletzt abgerufen 11.11.2013).

Dickmann, Frank und Ulrich Sax, "Konzept eines Betriebsmodells für Virtuelle Forschungsumgebungen." Version 14, 2010.
http://www.wissgrid.de/publikationen/deliverables/wp1/WissGrid_AP1_D1-4_final_v2.pdf (zuletzt abgerufen 11.11.2013).

- DFG – Deutsche Forschungsgemeinschaft. *Sicherung guter wissenschaftlicher Praxis*. Weinheim: WILEY-VCH Verlag GmbH, 1998.
http://www.dfg.de/download/pdf/dfg_im_profil/reden_stellungnahmen/download/empfehlung_wiss_praxis_0198.pdf (zuletzt abgerufen 11.11.2013).
- DINI -Deutsche Initiative für Netzwerkinformation e.V. "Arbeitsgruppe Virtuelle Forschungsumgebungen (vForum)." <http://www.dini.de/ag/vforum/> (zuletzt abgerufen 11.11.2013)
- DHd - Digital Humanities im deutschsprachigen Raum. "Digitale Geisteswissenschaften. Was sind die digitalen Geisteswissenschaften?" <http://www.dig-hum.de/digitale-geisteswissenschaften> (zuletzt abgerufen 11.11.2013).
- Enquete-Kommission „Internet und digitale Gesellschaft.“ *Sechster Zwischenbericht. Drucksache 17/12029*. 08. 01. 2013.
<http://dipbt.bundestag.de/dip21/btd/17/120/1712029.pdf> (zuletzt abgerufen 11.11.2013).
- ESFRI - European Strategy Forum on Research Infrastructures. "Roadmap 2010. Strategy Report on Research Infrastructures."
http://ec.europa.eu/research/infrastructures/pdf/esfri-strategy_report_and_roadmap.pdf (zuletzt abgerufen 11.11.2013).
- Fitzpatrick, Kathleen. "The Humanities, Done Digitally." *Debates in the digital humanities*, herausgegeben von Matthew K. Gold. Minneapolis: University of Minnesota Press, 2012. <http://dhdebates.gc.cuny.edu/debates/text/30> (zuletzt abgerufen 11.11.2013).
- Foster, Nancy Fried und Susan Gibbons. "Understanding Faculty to Improve Content Recruitment for Institutional Repositories." *D-Lib Magazine*, Vol. 11, No. 1 (January 2005). <http://www.dlib.org/dlib/january05/foster/01foster.html> (zuletzt abgerufen 11.11.2013).
- Funk, Stefan E. "7.6 Digitale Objekte und Formate." *nestor-Handbuch. Eine kleine Enzyklopädie der digitalen Langzeitarchivierung*. Version 2.3., herausgegeben von Heike Neuroth, Achim Oßwald, Regine Scheffel, Stefan Strathmann und Karsten Huth, 2010. <http://nestor.sub.uni-goettingen.de/handbuch/nestor-handbuch.pdf> (zuletzt abgerufen 11.11.2013).
- Funk, Stefan E. "8.3 Migration." *nestor-Handbuch. Eine kleine Enzyklopädie der digitalen Langzeitarchivierung*. Version 2.3., herausgegeben von Heike Neuroth, Achim Oßwald, Regine Scheffel, Stefan Strathmann und Karsten Huth, 2010.
<http://nestor.sub.uni-goettingen.de/handbuch/nestor-handbuch.pdf> (zuletzt

abgerufen 11.11.2013).

Given, Lisa. "Qualitative research in evidence-based practice: a valuable partnership." *Library Hi Tech*, Vol. 24, No.3 (2006): 376-386.
<http://dx.doi.org/10.1108/07378830610692145> (zuletzt abgerufen 11.11.2013).

Goldman, Nick, Paul Bertone, Siyuan Chen, Christophe Dessimoz, Emily M. LeProust, Botond Sipos und Ewan Birney. "Towards practical, high-capacity, low-maintenance information storage in synthesized DNA." *Nature*, Vol. 494, No. 7435. (23 January 2013): 77-80. <http://dx.doi.org/10.1038/nature11875> (zuletzt abgerufen 11.11.2013).

Golumbia, David. "Digital Humanities: Two Definitions." Article posted on January 20, 2013. *Uncomputing Blog*. <http://www.uncomputing.org/?p=203> (zuletzt abgerufen 11.11.2013).

Harvey, Ross. *Preserving Digital Materials*. 2nd. ed. Berlin: De Gruyter Saur, 2012.

Heiervang, Einar und R. Goodman. "Advantages and limitations of web-based survey: evidence from child mental health survey." *Social Psychiatry and Psychiatric Epidemiology*, Vol. 46, No. 1 (January 2011): 69-76.
<http://search.proquest.com/docview/845800833?accountid=14632> (zuletzt abgerufen 11.11.2013)

Higgins, Sarah. "What are Metadata Standards," 2007.
<http://www.dcc.ac.uk/resources/briefing-papers/standards-watch-papers/what-are-metadata-standards> (zuletzt abgerufen 11.11.2013).

High Level Expert Group on Scientific Data. "Riding the wave. How Europe can gain from the rising tide of scientific data. Final report of the High Level Expert Group on Scientific Data. A submission to the European Commission," 2010.
<http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/hlg-sdi-report.pdf> (zuletzt abgerufen 11.11.2013).

Hinton, Perry R. *Statistics Explained: A Guide for Social Science Students*. New York: Routledge, 1995.

Hockey, Susan. "The History of Humanities Computing." *A Companion to Digital Humanities*, herausgegeben von Susan Schreibman, Ray Siemens und John Unsworth. Oxford: Blackwell, 2004. <http://www.digitalhumanities.org/companion/> (zuletzt abgerufen 11.11.2013)

Huschka, Denis, Claudia Oellers, Notburga Ott und Gerd G. Wagner. "Datenmanagement und Data Sharing: Erfahrungen in den Sozial- und Wirtschaftswissenschaften." *Handbuch Forschungsdatenmanagement*, herausgegeben von Stephan Büttner, Hans-Christoph Hobohm und Lars Müller. Bad Honnef: Bock + Herchen Verlag, 2011: 35-49.

Jensen, Uwe; Alexia Katsanidou und Wolfgang Zenk-Möltgen. "Metadaten und Standards." *Handbuch Forschungsdatenmanagement*, herausgegeben von Stephan Büttner, Hans-Christoph Hobohm und Lars Müller, Bad Honnef: Bock + Herchen Verlag, 2011: 83-100.

Kelle, Udo. *Die Integration qualitativer und quantitativer Methoden in der empirischen Sozialforschung*. Wiesbaden: VS Verlag für Sozialwissenschaften, 2008.
<http://dx.doi.org/10.1007/978-3-531-91174-8> (zuletzt abgerufen 11.11.2013).

KIM – Kompetenzzentrum Interoperable Metadaten. "Forschungsdaten"
<https://wiki.dnb.de/display/DINIAGKIM/Forschungsdaten> (zuletzt abgerufen 11.11.2013).

King's College London. "Launch of the PERICLES project," gepostet am 22.02.2013.
<http://www.kcl.ac.uk/artshums/depts/ddh/newsrecords/pericles.aspx> (zuletzt abgerufen 11.11.2013).

Kirschenbaum, Matthew. "What Is Digital Humanities and What's It Doing in English Departments?" *ADE Bulletin*, 150 (2010): 1-7.
<http://mkirschenbaum.files.wordpress.com/2011/03/ade-final.pdf> (zuletzt abgerufen 11.11.2013).

KoLaWiss – Kooperative Langzeitarchivierung für Wissenschaftsstandorte.
"Zusammenfassung." <http://kolawiss.uni-goettingen.de/> (zuletzt abgerufen 11.11.2013).

Kommission Zukunft der Informationsinfrastruktur. "Gesamtkonzept für die Informationsinfrastruktur in Deutschland," 2011.
http://www.leibnizgemeinschaft.de/fileadmin/user_upload/downloads/Infrastruktur/KII_Gesamtkonzept.pdf (zuletzt abgerufen 11.11.2013).

Kromrey, Helmut. *Empirische Sozialforschung*. 12th ed. Stuttgart: Lucius & Lucius, 2009.

Kuipers, Tom und Jeffrey van der Hoeven. "PARSE.Insight: Insight into digital preservation of research output in Europe. Survey Report," 2009. http://www.parse-insight.eu/downloads/PARSE-Insight_D3-4_SurveyReport_final_hq.pdf (zuletzt

abgerufen 11.11.2013).

Leech, Nancy L., Anthony J. Onwuegbuzie. "A typology of mixed methods research designs." *Quality and Quantity*, Vol. 43, No. 2 (March 2009): 265-275.
<http://dx.doi.org/10.1007/s11135-007-9105-3> (zuletzt abgerufen 11.11.2013).

Liegmann, Hans und Heike Neuroth. "1. Einführung." *nestor-Handbuch. Eine kleine Enzyklopädie der digitalen Langzeitarchivierung. Version 2.3.*, herausgegeben von Heike Neuroth, Achim Oßwald, Regine Scheffel, Stefan Strathmann und Karsten Huth, 2010. <http://nestor.sub.uni-goettingen.de/handbuch/nestor-handbuch.pdf> (zuletzt abgerufen 11.11.2013).

Likert-Skalierung, *Gabler Wirtschaftslexikon*,
<http://wirtschaftslexikon.gabler.de/Archiv/12660/likert-skalierung-v5.html>, (zuletzt abgerufen 11.11.2013).

Ludwig, Jens und Harry Enke, Herausgeber. *Leitfaden zum Forschungsdaten-Management. Handreichungen aus dem WissGrid-Projekt*. Glückstadt: Verlag Werner Hülsbusch, 2013. http://www.wissgrid.de/publikationen/Leitfaden_Data-Management-WissGrid.pdf (zuletzt abgerufen 11.11.2013).

nestor-Arbeitsgruppe OAIS-Übersetzung. "Referenzmodell für ein Offenes Archiv-
Informations-System - Deutsche Übersetzung," 2012. http://files.d-nb.de/nestor/materialien/nestor_mat_16.pdf (zuletzt abgerufen 11.11.2013).

OAIS-Referenzmodell. Grafik von Fachlabor-Gubler.ch <http://www.fachlabor-gubler.ch/de/beratung/unsere-normativen-grundlagen/das-oais-modell.html> (zuletzt abgerufen 11.11.2013)

OECD – Organisation für Economic Co-Operation and development. "OECD Principles and Guidelines for Access to Research Data from Public Funding," 2007.
<http://www.oecd.org/sti/sci-tech/38500813.pdf> (zuletzt abgerufen 11.11.2013).

Oßwald, Achim, Regine Scheffel und Heike Neuroth. "1. Langzeitarchivierung von Forschungsdaten. Einführende Überlegungen." *Langzeitarchivierung von Forschungsdaten. Eine Bestandsaufnahme*, herausgegeben von Heike Neuroth, Stefan Strathmann, Achim Oßwald, Regine Scheffel, Jens Klump und Jens Ludwig. Boizenburg: Verlag Werner Hülsbusch, 2012: 13-22.

Parry, Dave. "The Digital Humanities or a Digital Humanism." *A Companion to Digital Humanities*, herausgegeben von Susan Schreibman, Ray Siemens und John Unsworth. Oxford: Blackwell, 2004. <http://dhdebates.gc.cuny.edu/debates/text/24> (zuletzt

abgerufen 11.11.2013).

Pempe, Wolfgang. "7. Geisteswissenschaften." *Langzeitarchivierung von Forschungsdaten. Eine Bestandsaufnahme*. Herausgegeben von Heike Neuroth, Stefan Strathmann, Achim Oßwald, Regine Scheffel, Jens Klump und Jens Ludwig. Boizenburg: Verlag Werner Hülsbusch, 2012: 137-160.

Rümpel, Stefanie. "Der Lebenszyklus der Forschungsdaten." *Handbuch Forschungsdatenmanagement*, herausgegeben von Stephan Büttner, Hans-Christoph Hobohm und Lars Müller. Bad Honnef: Bock + Herchen Verlag, 2011: 25-35.

Schnell, Rainer. *Survey-Interviews. Methoden standardisierter Befragungen*. Wiesbaden: VS Verlag für Sozialwissenschaften, 2012.

Severiens, Thomas und Eberhard R. Hilf. "Langzeitarchivierung von Rohdaten. Studie zum Stand vorhandener Forschungsdaten und Rohdaten aus wissenschaftlichen Tätigkeiten: Erfordernisse und Eignung zur Archivierung bzw. Zurverfügungstellung in Deutschland (Primärdaten)," 2006.
http://files.dnb.de/nestor/materialien/nestor_mat_06.pdf (zuletzt abgerufen 11.11.2013).

Taporwiki. "How do you define Humanities Computing / Digital Humanities?"
http://tapor.ualberta.ca/taporwiki/index.php/How_do_you_define_Humanities_Computing/_Digital_Humanities%3F (zuletzt abgerufen 11.11.2013).

Tenopir, Carol, Suzie Allard, Kimberly Douglass, Arsev Umur Aydinoglu, Lei Wu, Eleanor Read, Maribeth Manoff und Mike Frame. "Data Sharing by Scientists: Practices and Perceptions." *PLoS ONE* Vol. 6, No. 6 (June 2011): 1-21.
<http://www.deepdyve.com/lp/pubmed-central/data-sharing-by-scientists-practices-and-perceptions-E3q8VztFal?articleList=%2Fsearch%3Fquery%3DData%2Bsharing%2Bby%2Bscientist> (zuletzt abgerufen 11.11.2013).

Terrell, Steven R. "Mixed-Methods Research Methodologies." *The Qualitative Report*, Vol.17, No.1 (January 2012): 254-280.
<http://search.proquest.com/docview/920733426?accountid=14632> (zuletzt abgerufen 11.11.2013).

TextGrid. "Architektur des TextGridRep."
<http://www.textgrid.de/ueber-textgrid/materialien/#c329> (zuletzt abgerufen 11.11.2013).

TextGrid. "Das Projekt." <http://www.textgrid.de/de/ueber-textgrid/projekt/> (zuletzt abgerufen 11.11.2013).

TextGrid. "Tools, Services, Ressourcen." <http://www.textgrid.de/ueber-textgrid/tools-services-ressourcen/> (zuletzt abgerufen 11.11.2013).

Thibodeau, Kenneth. "Overview of Technological Approaches to Digital Preservation and Challenges in Coming Years¹," 2002.
<http://www.clir.org/pubs/reports/pub107/thibodeau.html> (zuletzt abgerufen 11.11.2013).

Ulrich, Dagmar. "8.2 Bitstream Preservation." *nestor-Handbuch. Eine kleine Enzyklopädie der digitalen Langzeitarchivierung*. Version 2.3, herausgegeben von Heike Neuroth, Achim Oßwald, Regine Scheffel, Stefan Strathmann und Karsten Huth, 2010.
<http://nestor.sub.uni-goettingen.de/handbuch/nestor-handbuch.pdf> (zuletzt abgerufen 11.11.2013).

Wissenschaftsrat. "Empfehlungen zu den Forschungsinfrastrukturen in den Geistes- und Sozialwissenschaften," 2011.
<http://www.wissenschaftsrat.de/download/archiv/10465-11.pdf> (zuletzt abgerufen 11.11.2013).

Zimmermann, Hans-Dieter und Joachim Pfister. "Bedarfsanalyse für ein Angebot „Digitale Langzeitarchivierung“ in den Geisteswissenschaften (data repository). Version 1.0," 2008.
http://www.sagw.ch/dms/sagw/laufende_projekte/DDZ/Beilagen/Auswertung.pdf (zuletzt abgerufen 11.11.2013).

10. Anhang

10.1 Fragebogen Online-Ansicht

Frame 1: Sicherheitsfrage

Langzeitarchivierung innerhalb virtueller Forschungsumgebungen im Bereich Digital Humanities

HUMBOLDT-UNIVERSITÄT ZU BERLIN

Es wird hauptsächlich eine Likert Skala zur Messung der persönlichen Einstellung verwendet. Die Antworten rangieren von JA (klare Zustimmung/Bejahung), EHER JA (teilweise Bejahung), WEDER NOCH (neutrale/keine Antwort), EHER NEIN (Teilweise Verneinung) bis zu NEIN (klare Ablehnung/Verneinung).

Bitte bestätigen Sie den Zugang zur Umfrage indem Sie die Sicherheitsfrage unten beantworten und auf Weiter klicken.

Sicherheitsfrage

Frame 2: Willkommenstext und Beschreibung

Langzeitarchivierung innerhalb virtueller Forschungsumgebungen im Bereich Digital Humanities

HUMBOLDT-UNIVERSITÄT ZU BERLIN

Es wird hauptsächlich eine Likert Skala zur Messung der persönlichen Einstellung verwendet. Die Antworten rangieren von JA (klare Zustimmung/Bejahung), EHER JA (teilweise Bejahung), WEDER NOCH (neutrale/keine Antwort), EHER NEIN (Teilweise Verneinung) bis zu NEIN (klare Ablehnung/Verneinung).

Untersucht wird die Frage, welche Möglichkeiten zur langfristigen Archivierung von Forschungsdaten an digitalen Arbeitsplätzen derzeit genutzt werden und welchen Bedarf die aktiv forschenden Wissenschaftlerinnen/Wissenschaftler in Zukunft hierfür sehen. Ich möchte Sie bitten so freundlich zu sein, sich ca. 10 Minuten Zeit zu nehmen, um die folgenden Fragen zu diesem Thema zu beantworten. Ihre Daten werden zu jeder Zeit streng anonym behandelt. Die erhobenen Daten werden nur im Kontext meiner Masterarbeit veröffentlicht werden. Ich bedanke mich herzlich für Ihre Kooperation!
Diese Umfrage enthält 20 Fragen.

Eine Bemerkung zum Datenschutz

Dies ist eine anonyme Umfrage.

Die Daten mit Ihren Antworten enthalten keinerlei auf Sie zurückzuführende/identifizierende Informationen, es sei denn bestimmte Fragen haben Sie explizit danach gefragt. Wenn Sie für diese Umfrage einen Zugangsschlüssel benutzt haben, so können Sie sicher sein, dass der Zugangsschlüssel nicht zusammen mit den Daten abgespeichert wurde. Er wird in einer getrennten Datenbank aufbewahrt und nur aktualisiert, um zu speichern, ob Sie diese Umfrage abgeschlossen haben oder nicht. Es gibt keinen Weg die Zugangsschlüssel mit den Umfrageergebnissen zusammenzuführen.

Frame 3: Fragebogen

Langzeitarchivierung innerhalb virtueller Forschungsumgebungen im Bereich Digital Humanities

HUMBOLDT-UNIVERSITÄT ZU BERLIN



Es wird hauptsächlich eine Likert Skala zur Messung der persönlichen Einstellung verwendet. Die Antworten rangieren von JA (klare Zustimmung/Bejahung), EHER JA (teilweise Bejahung), WEDER NOCH (neutrale/keine Antwort), EHER NEIN (Teilweise Verneinung) bis zu NEIN (klare Ablehnung/Verneinung).

0% 100%

* 1 Bearbeiten Sie digitale Forschungsmaterialien?

☐ Ja ☐ Nein

* 2 Welcher Art sind Ihre digitalen Forschungsmaterialien?

Bitte wählen Sie einen oder mehrere Punkte aus der Liste aus.

- ☐ Materialien digitalen Ursprungs (also genuin am Computer erstellt, z.B. ein Word-Dokument, Digitalfoto, ...)
- ☐ Digitalisate (z.B. von Dokumenten, Bildern, Filmen, Tonaufnahmen...)
- ☐ Mit OCR (Texterkennung) bearbeitete Textdateien (d.h. der Text ist durchsuchbar)
- ☐ Datenbanken
- ☐ Sonstiges

* 3 Welche Dateiformate bearbeiten Sie?

Bitte wählen Sie einen oder mehrere Punkte aus der Liste aus.

- ☐ Digitale Bilder von Text (z.B. digitalisierte Schriftstücke, Bücher, ...)
- ☐ Digitale Bilder (z.B. von Fotos, Gemälden, Zeichnungen, Graphiken, ...)
- ☐ Durchsuchbare Textdateien (z.B. Word, PDF, XML, LaTeX, ...)
- ☐ Tondateien (z.B. wav, aup, mp3, Ogg Vorbis, ...)
- ☐ Videodateien (z.B. wmv, mpeg I u. II, mp4, avi, mov, flv, swf, ...)
- ☐ Sonstiges

* 4 Verzeichnen Sie Metadaten (Eckdaten – Daten über Daten) zu Ihren digitalen Forschungsdateien?

Bitte wählen Sie eine der folgenden Antworten:

- ☒ ja
- ☐ eher ja
- ☐ weder noch
- ☐ eher nein
- ☐ nein

* 4a Sind dies nach anerkannten Standards vergebene Metadaten (z.B. Dublin Core, RDF/ Resource Description Framework, ...)?

Bitte wählen Sie eine der folgenden Antworten:

- ☐ ja
- ☐ eher ja
- ☐ weder noch
- ☐ eher nein
- ☐ nein

* 5 Denken Sie, dass es für Ihre Arbeit wichtig ist Metadaten zu verzeichnen?
Bitte wählen Sie eine der folgenden Antworten:

- ☐ ja
- ☐ eher ja
- ☐ weder noch
- ☐ eher nein
- ☐ nein

* 6 Denken Sie, dass Sie Ihre digitalen Dateien längerfristig, d.h. in einigen Jahren noch benötigen werden?
Bitte wählen Sie eine der folgenden Antworten:

- ☐ ja
- ☐ eher ja
- ☐ weder noch
- ☒ eher nein
- ☐ nein

* 7 Ist bei Ihnen schon das Problem aufgetreten, dass Sie nicht mehr auf digitale Forschungsmaterialien zugreifen konnten, die Sie noch brauchten?
Bitte wählen Sie eine der folgenden Antworten:

- ☐ ja
- ☐ eher ja
- ☐ weder noch
- ☐ eher nein
- ☐ nein

* 8 Erfolgt derzeit eine Sicherheitskopie (Backup) Ihrer digitalen Daten?
Bitte wählen Sie eine der folgenden Antworten:

- ☐ ja
- ☐ eher ja
- ☐ weder noch
- ☐ eher nein
- ☐ nein

* 9 Würden Sie sich wünschen, dass Ihre Forschungsdaten auf unbestimmte Zeit sicher von Ihrer Universität archiviert werden?
Bitte wählen Sie eine der folgenden Antworten:

- ☐ ja
- ☐ eher ja
- ☐ weder noch
- ☐ eher nein
- ☐ nein

* 10 Würden Sie Ihre Daten einer wissenschaftlichen Institution, die als externer Dienstleister auftritt, zum Speichern anvertrauen?
Bitte wählen Sie eine der folgenden Antworten:

- ☐ ja
- ☐ eher ja
- ☐ weder noch
- ☐ eher nein
- ☐ nein

* **11** Denken Sie, dass Dritte (z.B. andere Wissenschaftler) Ihre digitalen Forschungsdateien zukünftig als hilfreich für weitere Forschungsprojekte erachten könnten?
Bitte wählen Sie eine der folgenden Antworten:

- ☐ ja
- ☐ eher ja
- ☐ weder noch
- ☐ eher nein
- ☐ nein

* **12** Würden Sie Ihre digitalen Forschungsdateien mit anderen Wissenschaftlern teilen?
Bitte wählen Sie eine der folgenden Antworten:

- ☐ ja
- ☐ eher ja
- ☐ weder noch
- ☐ eher nein
- ☐ nein

* **Würden Sie Ihre digitalen Forschungsdateien mit anderen teilen, wenn eine Rechtemanagement den Zugriff regeln würde?**
Bitte wählen Sie eine der folgenden Antworten:

- ☒ ja
- ☐ eher ja
- ☐ weder noch
- ☐ eher nein
- ☐ nein

* **12a** Würden Sie ihre digitalen Forschungsdateien mit anderen Teilen, wenn ein Rechtemanagement den Zugriff regeln würde?
Bitte wählen Sie eine der folgenden Antworten:

- ☐ ja
- ☐ eher ja
- ☐ weder noch
- ☐ eher nein
- ☐ nein

* **13** Teilen Sie vielleicht bereits jetzt schon Ihre digitalen Forschungsdateien mit anderen Personen?
Bitte wählen Sie eine der folgenden Antworten:

- ☐ ja
- ☐ eher ja
- ☐ weder noch
- ☐ eher nein
- ☐ nein

* **13a** Wie teilen sie Ihre digitalen Forschungsdateien mit anderen?
Bitte wählen Sie einen oder mehrere Punkte aus der Liste aus.

- ☐ via E-Mail
- ☐ Speichermedien (z.B. CD-ROM, USB-Stick, etc.)
- ☐ gemeinsames Netzlaufwerk
- ☐ über ein Fachportal
- ☐ über einen webbasierten gemeinsamen „Raum“ (Wiki, Blog, Google Drive, Clouds, etc.)
- ☐ über einen Verlag
- ☐ Sonstiges

• **14** Würden Sie Ihre digitalen Daten mit anderen Wissenschaftlern gemeinsam beforschen?
Bitte wählen Sie eine der folgenden Antworten:

- ☐ ja
- ☐ eher ja
- ☐ weder noch
- ☐ eher nein
- ☐ nein

• **15** Zum Schluss möchte ich Sie bitten noch einige Angaben zu statistischen Zwecken zu machen: Geschlecht?

- ☐ weiblich
- ☐ männlich

• **16** Alter?
Bitte wählen Sie eine der folgenden Antworten:

- ☐ 20 - 29
- ☐ 30 - 39
- ☐ 40 - 49
- ☐ 50 - 59
- ☐ 60 - 69

• **17** Akademischer Grad?
Bitte wählen Sie eine der folgenden Antworten:

- ☐ Dr. habil.
- ☐ Juniorprof.
- ☐ PD Dr.
- ☐ PD Dr. habil.
- ☐ apl. Prof.Dr.
- ☐ Prof. Dr.
- ☐ Sonstiges

Später fortfahren

Absenden

Umfrage verlassen und Antworten löschen

10.2 Umfrage Logikdatei

Keine Syntax-Fehler in dieser Umfrage gefunden

Logikdatei für die Umfrage #[42924]: Langzeitarchivierung innerhalb virtueller Forschungsumgebungen im Bereich Digital Humanities

Beschreibung: Es wird hauptsächlich eine Likert Skala zur Messung der persönlichen Einstellung verwendet. Die Antworten rangieren von JA (klare Zustimmung/Bejahung), EHER JA (teilweise Bejahung), WEDER NOCH (neutrale/keine Antwort), EHER NEIN (teilweise Verneinung) bis zu NEIN (klare Ablehnung/Verneinung).

Willkommenstext: Untersucht wird die Frage, welche Möglichkeiten zur langfristigen Archivierung von Forschungsdaten an digitalen Arbeitsplätzen

derzeit genutzt werden und welchen Bedarf die aktiv forschenden Wissenschaftlerinnen/Wissenschaftler in Zukunft hierfür sehen. Ich möchte Sie bitten so freundlich zu sein, sich ca. 10 Minuten Zeit zu nehmen, um die folgenden Fragen zu diesem Thema zu beantworten. Ihre Daten werden zu jeder Zeit streng anonym behandelt. Die erhobenen Daten werden nur im Kontext meiner Masterarbeit veröffentlicht werden. Ich bedanke mich herzlich für Ihre Zeit und Kooperation!

Endnachricht:

Herzlichen Dank für Ihre Zeit und Unterstützung!

#	Name [ID]	Relevanz [Validierung] (Standard)	Text [Hilfe] (Hinweis)
G-0	Alle Fragen [GID 1576]	1	
Q-0	*1 [QID 18277] Ja/Nein [Y]	1	Bearbeiten Sie digitale Forschungsmaterialien?
Q-1	*2 [QID 18278] Mehrfachauswahl [M]	1	Welcher Art sind Ihre digitalen Forschungsmaterialien? Fragenattribut Wert assessment_value 1 display_columns 1
SQ-1	2_SQ001		Materialien digitalen Ursprungs (also genuin am Computer erstellt, z.B. ein Word- Dokument, Digitalfoto, ...)
SQ-2	2_SQ002		Digitalisate (z.B. von Dokumenten, Bildern, Filmen, Tonaufnahmen...)
SQ-3	2_SQ003		Mit OCR (Texterkennung) bearbeitete Textdateien (d.h. der Text ist durchsuchbar)
SQ-4	2_SQ004		Datenbanken
SQ-5	2_SQ005		Sonstiges
Q-2	*3 [QID 18279] Mehrfachauswahl [M]	1	Welche Dateiformate bearbeiten Sie? Fragenattribut Wert assessment_value 1 display_columns 1
SQ-1	3_SQ001		Digitale Bilder von Text (z.B. digitalisierte Schriftstücke, Bücher, ...)
SQ-2	3_SQ002		Digitale Bilder (z.B. von Fotos, Gemälden, Zeichnungen, Graphiken, ...)
SQ-3	3_SQ003		Durchsuchbare Textdateien (z.B. Word, PDF, XML, LaTeX, ...)
SQ-4	3_SQ004		Tondateien (z.B. wav, aup, mp3, Ogg Vorbis, ...)
SQ-5	3_SQ005		Videodateien (z.B. wmv, mpeg I u. II, mp4, avi, mov, flv, swf, ...)

SQ-6 3_SQ006

<p>*4 [QID 18280]</p>		<p>Sonstiges</p>	
Q-3	Liste (Optionsfelder) [L]	1 (Voreinstellung: A1)	Verzeichnen Sie Metadaten (Eckdaten – Daten über Daten) zu Ihren digitalen Forschungsdateien?
		Fragenattribut	Wert
		display_columns	1
A[0]-1	A1	[VALUE: A1]	ja
A[0]-2	A2	[VALUE: A2]	eher ja
A[0]-3	A3	[VALUE: A3]	weder noch
A[0]-4	A4	[VALUE: A4]	eher nein
A[0]-5	A5	[VALUE: A5]	nein
<p>*4a [QID 18281]</p>		<p>Sind dies nach anerkannten Standards vergebene Metadaten (z.B. Dublin Core, RDF/ Resource Description Framework, ...)?</p>	
Q-4	Liste (Optionsfelder) [L]	((4.NAOK == "A1" or 4.NAOK == "A2"))	
		Fragenattribut	Wert
		display_columns	1
A[0]-1	A1	[VALUE: A1]	ja
A[0]-2	A2	[VALUE: A2]	eher ja
A[0]-3	A3	[VALUE: A3]	weder noch
A[0]-4	A4	[VALUE: A4]	eher nein
A[0]-5	A5	[VALUE: A5]	nein
<p>*5 [QID 18282]</p>		<p>Denken Sie, dass es für Ihre Arbeit wichtig ist Metadaten zu verzeichnen?</p>	
Q-5	Liste (Optionsfelder) [L]	1	
		Fragenattribut	Wert
		display_columns	1
A[0]-1	A1	[VALUE: A1]	ja
A[0]-2	A2	[VALUE: A2]	eher ja
A[0]-3	A3	[VALUE: A3]	weder noch
A[0]-4	A4	[VALUE: A4]	eher nein
A[0]-5	A5	[VALUE: A5]	nein

*6 [QID 18283]		Denken Sie, dass Sie Ihre digitalen Dateien längerfristig, d.h. in einigen Jahren noch benötigen werden?	
Q-6	Liste (Optionsfelder) [L]	1	Fragenattribut Wert display_columns 1
A[0]-1	A1	[VALUE: A1]	ja
A[0]-2	A2	[VALUE: A2]	eher ja
A[0]-3	A3	[VALUE: A3]	weder noch
A[0]-4	A4	[VALUE: A4]	eher nein
A[0]-5	A5	[VALUE: A5]	nein
*7 [QID 18284]		Ist bei Ihnen schon das Problem aufgetreten, dass Sie nicht mehr auf digitale Forschungsmaterialien zugreifen konnten, die Sie noch brauchten?	
Q-7	Liste (Optionsfelder) [L]	1	Fragenattribut Wert display_columns 1
A[0]-1	A1	[VALUE: A1]	ja
A[0]-2	A2	[VALUE: A2]	eher ja
A[0]-3	A3	[VALUE: A3]	weder noch
A[0]-4	A4	[VALUE: A4]	eher nein
A[0]-5	A5	[VALUE: A5]	nein
*8 [QID 18285]		Erfolgt derzeit eine Sicherheitskopie (Backup) Ihrer digitalen Daten?	
Q-8	Liste (Optionsfelder) [L]	1	Fragenattribut Wert display_columns 1
A[0]-1	A1	[VALUE: A1]	ja
A[0]-2	A2	[VALUE: A2]	eher ja
A[0]-3	A3	[VALUE: A3]	weder noch
A[0]-4	A4	[VALUE: A4]	eher nein

A[0]- 5	A5	[VALUE: A5]	nein
*9 [QID 18286]			
Q-9	Liste (Optionsfelder) [L]	1	Würden Sie sich wünschen, dass Ihre Forschungsdaten auf unbestimmte Zeit sicher von Ihrer Universität archiviert werden? Fragenattribut Wert display_columns 1
A[0]- 1	A1	[VALUE: A1]	ja
A[0]- 2	A2	[VALUE: A2]	eher ja
A[0]- 3	A3	[VALUE: A3]	weder noch
A[0]- 4	A4	[VALUE: A4]	eher nein
A[0]- 5	A5	[VALUE: A5]	nein
*10 [QID 18287]			
Q-10	Liste (Optionsfelder) [L]	1	Würden Sie Ihre Daten einer wissenschaftlichen Institution, die als externer Dienstleister auftritt, zum Speichern anvertrauen? Fragenattribut Wert display_columns 1
A[0]- 1	A1	[VALUE: A1]	ja
A[0]- 2	A2	[VALUE: A2]	eher ja
A[0]- 3	A3	[VALUE: A3]	weder noch
A[0]- 4	A4	[VALUE: A4]	eher nein
A[0]- 5	A5	[VALUE: A5]	nein
*11 [QID 18288]			
Q-11	Liste (Optionsfelder) [L]	1	Denken Sie, dass Dritte (z.B. andere Wissenschaftler) Ihre digitalen Forschungsdateien zukünftig als hilfreich für weitere Forschungsprojekte erachten könnten? Fragenattribut Wert display_columns 1
A[0]- 1	A1	[VALUE: A1]	ja
A[0]- 2	A2	[VALUE: A2]	eher ja
A[0]-	A3	[VALUE: A3]	weder noch

3

A[0]-
4 **A4** [VALUE: A4] eher nein

A[0]-
5 **A5** [VALUE: A5] nein

***12**

[\[QID 18289\]](#)

Q-12 Liste
(Optionsfelder)
[L]

1

Würden Sie Ihre digitalen Forschungsdateien mit anderen Wissenschaftlern teilen?

Fragenattribut Wert

display_columns 1

A[0]-
1 **A1** [VALUE: A1] ja

A[0]-
2 **A2** [VALUE: A2] eher ja

A[0]-
3 **A3** [VALUE: A3] weder noch

A[0]-
4 **A4** [VALUE: A4] eher nein

A[0]-
5 **A5** [VALUE: A5] nein

***12a**

[\[QID 18290\]](#)

Q-13 Liste
(Optionsfelder)
[L]

((12.NAOK == "A4"
or 12.NAOK ==
"A5"))

Würden Sie ihre digitalen Forschungsdateien mit anderen Teilen, wenn ein Rechtemanagement den Zugriff regeln würde?

Fragenattribut Wert

display_columns 1

A[0]-
1 **A1** [VALUE: A1] ja

A[0]-
2 **A2** [VALUE: A2] eher ja

A[0]-
3 **A3** [VALUE: A3] weder noch

A[0]-
4 **A4** [VALUE: A4] eher nein

A[0]-
5 **A5** [VALUE: A5] nein

***13**

[\[QID 18291\]](#)

Q-14 Liste
(Optionsfelder)
[L]

1

Teilen Sie vielleicht bereits jetzt schon Ihre digitalen Forschungsdateien mit anderen Personen?

Fragenattribut Wert

display_columns 1

A[0]-
1 **A1** [VALUE: A1] ja

A[0]-
2 **A2** [VALUE: A2] eher ja

A[0]- 3	A3	[VALUE: A3]	weder noch
A[0]- 4	A4	[VALUE: A4]	eher nein
A[0]- 5	A5	[VALUE: A5]	nein
Q-15	*13a [QID 18292] Mehrfachauswahl [M]	((13.NAOK == "A1" or 13.NAOK == "A2"))	Wie teilen sie Ihre digitalen Forschungsdateien mit anderen? Fragenattribut Wert assessment_value 1 display_columns 1 via E-Mail Speichermedien (z.B. CD-ROM, USB-Stick, etc.) gemeinsames Netzlaufwerk über ein Fachportal über einen webbasierten gemeinsamem „Raum“ (Wiki, Blog, Google Drive, Clouds, etc.) über einen Verlag Sonstiges
SQ-1	13a_SQ001		
SQ-2	13a_SQ002		
SQ-3	13a_SQ003		
SQ-4	13a_SQ004		
SQ-5	13a_SQ005		
SQ-6	13a_SQ006		
SQ-7	13a_SQ007		
Q-16	*14 [QID 18293] Liste (Optionsfelder) [L]	1	Würden Sie Ihre digitalen Daten mit anderen Wissenschaftlern gemeinsam beforschen? Fragenattribut Wert display_columns 1
A[0]- 1	A1	[VALUE: A1]	ja
A[0]- 2	A2	[VALUE: A2]	eher ja
A[0]- 3	A3	[VALUE: A3]	weder noch
A[0]- 4	A4	[VALUE: A4]	eher nein
A[0]- 5	A5	[VALUE: A5]	nein
Q-17	*15 [QID 18294] Geschlecht [G]	1	Zum Schluss möchte ich Sie bitten noch einige Angaben zu statistischen Zwecken zu machen: Geschlecht?
Q-18	*16 [QID 18295] Liste (Optionsfelder) [L]	1	Alter? Fragenattribut Wert display_columns 1
A[0]-	A1	[VALUE: A1]	20 - 29

1			
A[0]- 2	A2	[VALUE: A2]	30 - 39
A[0]- 3	A3	[VALUE: A3]	40 - 49
A[0]- 4	A4	[VALUE: A4]	50 - 59
A[0]- 5	A5	[VALUE: A5]	60 - 69
*17			
[QID 18296]			
Q-19	Liste (Optionsfelder) [L]	1	Akademischer Grad? Fragenattribut Wert display_columns 1
A[0]- 1	A1	[VALUE: A1]	Dr. habil.
A[0]- 2	A2	[VALUE: A2]	Juniorprof.
A[0]- 3	A3	[VALUE: A3]	PD Dr.
A[0]- 4	A4	[VALUE: A4]	PD Dr. habil.
A[0]- 5	A5	[VALUE: A5]	apl. Prof.Dr.
A[0]- 6	A6	[VALUE: A6]	Prof. Dr.
A[0]- 7	A7	[VALUE: A7]	Sonstiges